

# **Communications 1 [5ETC0]**

## **Course Reader**

**Prof. Dr. Oded Raz,**

**Nart Gashi,**

**Ion Kolkhuis Tanke**

# Contents

<b>1 Preface</b>	<b>1</b>
1.1 Course motivation . . . . .	1
1.2 Reader structure . . . . .	1
1.3 Installing Minilabs . . . . .	3
1.4 Instruction exercises . . . . .	3
<b>2 Orthogonal basis and the Fourier transform/series</b>	<b>4</b>
2.1 Learning objectives . . . . .	4
2.2 Motivation . . . . .	4
2.3 Intuition Behind the Fourier analysis . . . . .	5
2.3.1 Basis vectors and vector space . . . . .	5
2.3.2 Dot Product . . . . .	6
2.3.3 Orthogonal Basis . . . . .	6
2.3.4 Basis Functions . . . . .	7
2.3.5 Hilbert Space: Extending Vector Projections to Functions . . . . .	7
2.3.6 Orthogonal/Orthonormal Basis Functions . . . . .	8
2.4 Fourier Series . . . . .	10
2.4.1 Quadrature Form to Obtain Fourier Series Coefficients . . . . .	14
2.5 Fourier Transform . . . . .	16
2.5.1 Properties of the Fourier transform . . . . .	18
<b>3 Sampling theory</b>	<b>19</b>
3.1 Learning objectives . . . . .	19
3.2 Motivation . . . . .	19
3.3 Band-limited and time-limited signals . . . . .	20
3.3.1 Illustration of band-limited and time-limited signals . . . . .	20
3.4 Impulse sampling & Sampling theorem . . . . .	21
3.4.1 Signal reconstruction . . . . .	24
3.4.2 Aliasing . . . . .	26
3.5 Dimensionality theorem . . . . .	28
3.6 Instruction Exercises - Sampling Theory . . . . .	29
<b>4 Sampling Methods (Pulse Amplitude Modulation)</b>	<b>30</b>
4.1 Learning objectives . . . . .	30
4.2 Motivation . . . . .	30
4.3 Gating (natural sampling) . . . . .	31
4.3.1 Spectrum of natural sampling . . . . .	31
4.4 Flat-top sampling (or instantaneous sampling) . . . . .	33
4.4.1 Spectrum of flat-top sampling . . . . .	33
4.5 Instruction Exercises - Sampling methods . . . . .	37
<b>5 Digitization</b>	<b>38</b>
5.1 Learning objectives . . . . .	38
5.2 Motivation . . . . .	38
5.3 Pulse code modulation (PCM) . . . . .	38
5.3.1 From samples to quantized information . . . . .	38
5.3.2 Bit-rate $R$ . . . . .	39
5.3.3 Spectral efficiency $\eta$ . . . . .	41

5.3.4	PCM Bandwidth . . . . .	41
5.4	Signal-to-noise ratio ( <i>after signal reconstruction</i> ) . . . . .	42
5.4.1	Receiver output signal-to-noise ratio . . . . .	43
5.5	Bit-error probability $P_e$ and channel noise . . . . .	47
5.5.1	Probability and random variables . . . . .	47
5.5.2	Probability density functions (PDFs) . . . . .	48
5.5.3	Gaussian distribution (or normal distribution) . . . . .	49
5.5.4	Additive white Gaussian noise (AWGN) . . . . .	50
5.5.5	Q-function . . . . .	52
5.5.6	Computing Bit-error probability on AWGN channels . . . . .	53
5.5.7	Relating SNR <sub>in</sub> and SNR <sub>out</sub> . . . . .	55
5.6	Instruction Exercises — Digitization . . . . .	62
<b>6</b>	<b>Digital signaling</b>	<b>64</b>
6.1	Learning objectives . . . . .	64
6.2	Motivation . . . . .	64
6.3	Multi-level signaling . . . . .	64
6.3.1	Advantages and Disadvantages of multilevel signaling . . . . .	65
6.3.2	Bits per sample and bits per level, the difference . . . . .	67
6.4	Baud (or Symbol) rate . . . . .	68
6.4.1	Bandwidth estimation . . . . .	69
<b>7</b>	<b>Linecodes and their spectras</b>	<b>70</b>
7.1	Learning objectives . . . . .	70
7.2	Motivation . . . . .	70
7.3	Power spectral density (PSD) . . . . .	70
7.3.1	Derivation of PSD . . . . .	70
7.4	Unipolar NRZ . . . . .	72
7.5	Polar NRZ . . . . .	74
7.6	Unipolar RZ . . . . .	75
7.7	Bipolar RZ . . . . .	76
7.8	Manchester . . . . .	77
7.9	Power spectra for multilevel signaling . . . . .	78
7.10	Spectral efficiency of linecodes . . . . .	78
7.11	Instruction Exercises - Digital signaling and line codes . . . . .	80
<b>8</b>	<b>Inter-symbol interference</b>	<b>82</b>
8.1	Learning objectives . . . . .	82
8.2	Motivation . . . . .	82
8.3	What is inter-symbol interference? . . . . .	83
8.4	Modelling ISI . . . . .	85
8.5	Nyquist's first method (Zero ISI) . . . . .	86
8.6	Raised cosine-rolloff Nyquist filtering . . . . .	88
8.7	Instruction exercises - ISI . . . . .	93
<b>9</b>	<b>Information theory</b>	<b>96</b>
9.1	Learning objectives . . . . .	96
9.2	Motivation . . . . .	96
9.3	Shannon–Hartley channel capacity theorem . . . . .	97
9.4	Error detection and correction schemes . . . . .	99

9.4.1	Parity bit checking - (Error detection) . . . . .	99
9.4.2	Limitations of parity bit error detection . . . . .	100
9.4.3	Hamming codes - (Error correction and detection) . . . . .	102
9.5	Instruction exercises - Information Theory . . . . .	105
<b>10</b>	<b>Amplitude, Frequency and Phase Modulation</b>	<b>106</b>
10.1	Learning objectives . . . . .	106
10.2	Motivation . . . . .	106
10.3	Baseband vs Bandpass . . . . .	106
10.3.1	Up and down conversion . . . . .	107
10.4	Amplitude Modulation (AM) . . . . .	109
10.4.1	Percentage of modulation . . . . .	110
10.4.2	Modulation efficiency . . . . .	110
10.4.3	Peak envelope power . . . . .	110
10.4.4	Spectrum of AM signals . . . . .	111
10.5	Double-Sideband Suppressed Carrier (DSB-SC) Modulation . . . . .	112
10.5.1	Spectrum of DSB-SC signals . . . . .	113
10.6	AM Detector circuits . . . . .	114
10.6.1	Envelope Detector . . . . .	114
10.6.2	Product Detector . . . . .	115
10.7	Frequency and phase modulation . . . . .	117
10.7.1	Frequency and phase modulation index . . . . .	119
10.7.2	Carson's Rule . . . . .	120
10.8	Frequency Spectrum of FM and PM . . . . .	121
10.9	Instruction Exercises - AM/FM/PM . . . . .	124
<b>11</b>	<b>The Physical channel</b>	<b>128</b>
11.1	Learning objectives . . . . .	128
11.2	Motivation . . . . .	128
11.3	Wired and wireless comparison . . . . .	128
11.4	Wireless channel . . . . .	130
11.4.1	RF propagation . . . . .	130
11.4.2	Free space wireless propagation . . . . .	130
11.4.3	The case of a single reflection (Deriving the $d^{-4}$ rule) . . . . .	131
11.4.4	Knife Edge . . . . .	134
11.5	Wired Connections . . . . .	136
11.5.1	Introduction . . . . .	136
11.6	Exercises . . . . .	140
<b>12</b>	<b>Appendix</b>	<b>141</b>
12.1	Working with dB's . . . . .	141
<b>13</b>	<b>Physical channel equations</b>	<b>143</b>
<b>14</b>	<b>Appendix C: Autocorrelation and PSD relation</b>	<b>144</b>

# 1 Preface

## 1.1 Course motivation

In today's hyper-connected world, the very fabric of our society is interwoven with the intricate threads of telecommunications. From the moment we wake up to the second we fall asleep, we are immersed in a seamless flow of information, thanks to the wonders of modern communication technology. Have you ever wondered how your smartphone manages to connect you with friends and family across the globe in an instant? Or how the internet allows you to access a wealth of knowledge at your fingertips? These marvels are the result of the fascinating world of telecommunications.

The course, communication 1, will introduce the underlying basic concepts that allow modern society to transfer large amounts of information across very large distance with good fidelity. In Figure 1 we illustrate how analog information from our surrounding world (audio signal picked up by a microphone) goes through the process of digitization and all subsequent steps, to emerge again as an analog signal at the speaker on the other side. In this course, we will explain and offer tools for students to understand and implement the basic building blocks of such a system.

## 1.2 Reader structure

The reader follows the basic building blocks illustrated in Figure 1 with chapters dedicated to all major functions and concepts included in the process. In each relevant section we have also included instructions for the use of the mini-labs and relevant exercises students are invited to complete in order to master the topic. For all exercises included in the reader, annotated power point solutions are available as well as video's of the lecturer solving the problems with in depth explanation of the approach for the solution. It is recommended that students attempt to solve exercises first before watching the video solutions or reading the annotated solutions to enhance learning.

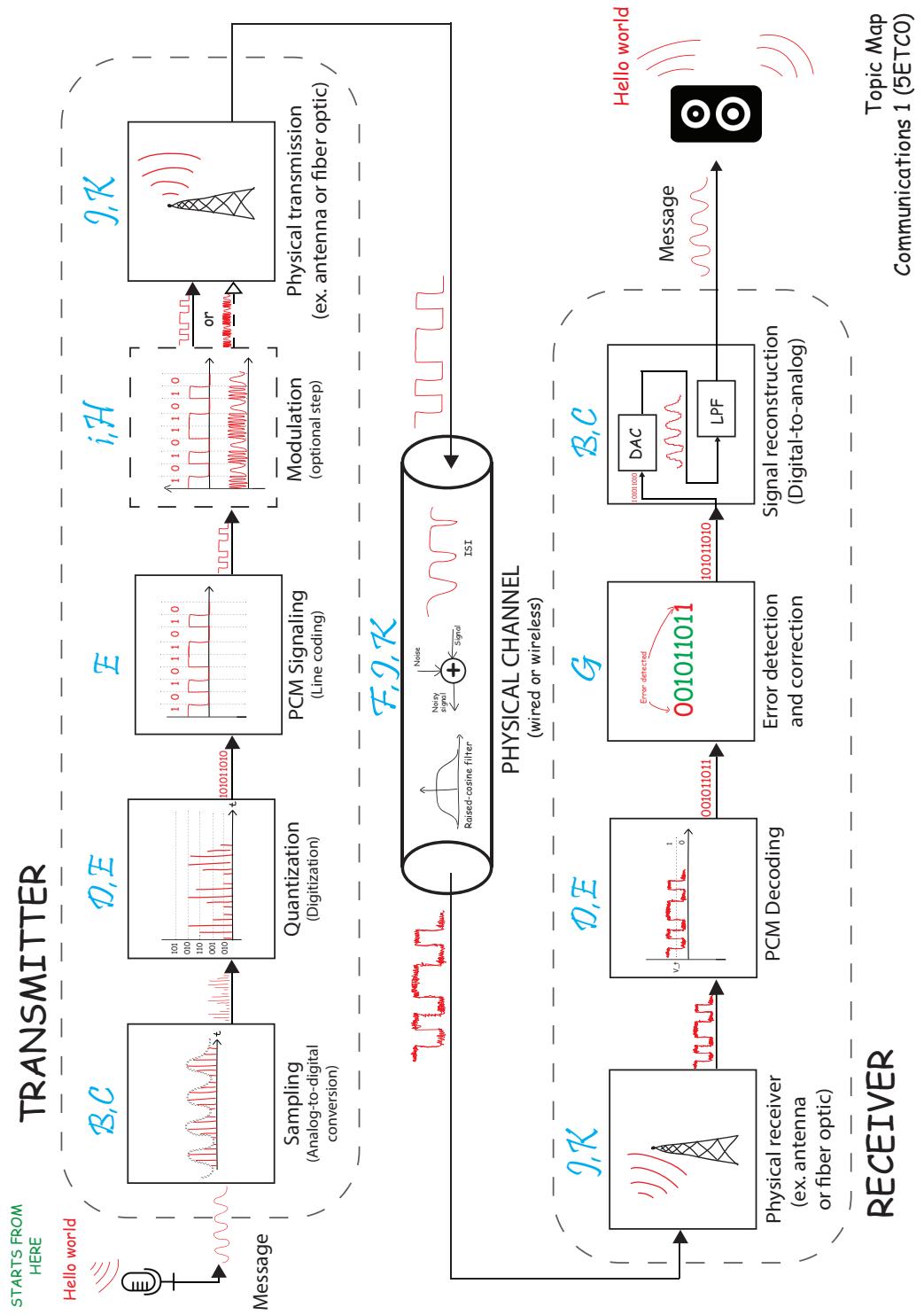


Figure 1: Communication 1 topic map. Observe how the message sent (colored in red), changes form during each step in the digital communication system. In blue, the respective modules in Canvas dealing with these topics

### 1.3 Installing Minilabs

Minilabs are MATLAB apps which will be used throughout the reader as a means to visualize and try out different concepts while you learn the topic. It is highly recommended to try out all minilab exercises, and ask the TA's on discord or during instruction hours regarding the answers to the minilab exercises.

The minilab exercises in the reader, will have a red box and be marked as a minilab exercise.

There are in total 7 minilabs, which cover almost all topics that are important for this course. These are,

- Minilab 1 - Fourier series
- Minilab 2 - Sampling and quantization
- Minilab 3 - Digital communication system (Digitization, line coding)
- Minilab 4 - Intersymbol interference
- Minilab 5 - Information theory
- Minilab 6 - Analog communication system (FM/AM modulation)
- Minilab 7 - Fiber optic channel

The relevant minilabs will be available to be downloaded under each module section in canvas.

The installation guide for a minilab may be found under **Modules → Minilab installation guide.**

### 1.4 Instruction exercises

At the end of each chapter there will be exercises for you to solve them (Excluding chapter 2 and 11). We highly recommend that you try them out yourself firstly, and then try to check the solution of those instruction.

Note that under each relevant module section, you will find a set of slides which has the solution of each instruction. Moreover, for most instruction there are videos of the teacher solving them and explaining them, which can also be found on each relevant module section.

## 2 Orthogonal basis and the Fourier transform/series

### 2.1 Learning objectives

Students completing this chapter should have learned:

1. How to represent a vector in N dimensions using a group of N orthogonal basis vectors.
2. Understand that functions can be represented by a set of other orthogonal functions.
3. Can prove that two trigonometric functions (sine or cosine) are orthogonal over a span T if their frequencies are harmonics of  $f=1/T$  (so  $n/T$  and  $k/T$  where  $n \neq k$ ).
4. Know that every periodic function can be represented by an infinite sum of sine and cosine functions - A Fourier series expansion.
5. Can calculate the Fourier coefficients and Fourier transform of a given function.
6. Understand that a Fourier series or Fourier transform are simply an alternative representation of a function and not a different function!

### 2.2 Motivation

The Fourier theorem is one of the most important tools in engineering, with use cases in all fields ranging from power electronics, signal processing, radio-astronomy, image processing and naturally, telecommunications. Without this mathematical tool many subsequent developments could have never been achieved, and without our understanding of it it will be very difficult to fully understand the concepts and ideas detailed in this course. For that reason, the concept of Fourier series is covered in early courses in the EE curriculum. Here we include a short overview of the main finding of Fourier theory related to time and frequency representations of signals and the difference between the Fourier transform and the Fourier series.

Originally devised to address heat transfer problems, Joseph Fourier proposed that any periodic function can be approximated by a finite sum of sine and cosine functions. Although this statement only holds under certain conditions, for functions that meet these conditions there is an alternative representation, known as the “Fourier series.” To obtain this series representation, one applies the “Fourier transform” to calculate the necessary coefficients, as will be discussed below.

## 2.3 Intuition Behind the Fourier analysis

Fourier analysis is a mathematical technique that decomposes a complex signal into its constituent frequencies, revealing the underlying sinusoidal components that make up the signal. This process provides an alternative representation of the signal in the frequency domain (referred to as spectrum).

In this subsection, we will develop a step-by-step intuitive understanding of the Fourier Analysis—why it works and how it works—beginning with fundamental concepts from vectors.

### 2.3.1 Basis vectors and vector space

Before talking about *basis vectors*, let's begin with the idea of a *vector space*. In simple terms, a vector space is any collection of vectors that can be added together and scaled by numbers (called *scalars*) in a way that follows certain consistent rules.

Suppose we have a two-dimensional plane; every point in the plane can be seen as a vector which can be composed of two vectors,  $\hat{i}$  which is a basis vector spanning x-coordinate direction and  $\hat{j}$  spanning y-coordinate direction (see figure below).

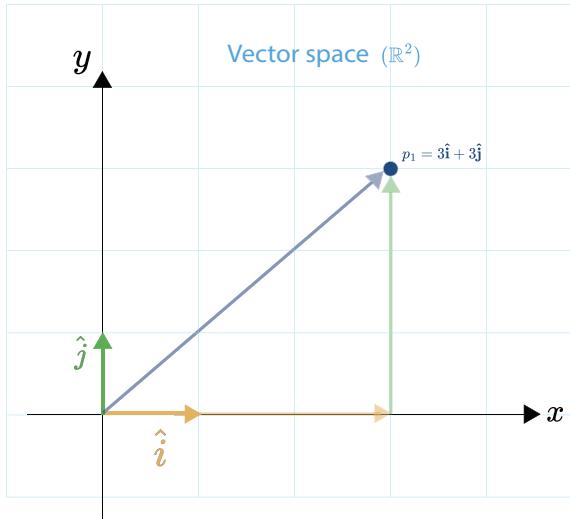


Figure 2: 2D ( $\mathbf{R}^2$ ) Vector space visualization with basis vectors  $\hat{i}, \hat{j}$ .

We can write down the bases vectors as follows

$$\hat{i} = (1, 0), \quad \hat{j} = (0, 1).$$

We can extend this formulation to higher-dimensions than 2,

$$R^2, \text{ defined by } \mathbf{i} = (1, 0), \mathbf{j} = (0, 1) \quad (1)$$

$$R^3, \text{ defined by } \mathbf{i} = (1, 0, 0), \mathbf{j} = (0, 1, 0), \mathbf{k} = (0, 0, 1) \quad (2)$$

$$R^n, \text{ defined by } \mathbf{i} = (1, 0, \dots, 0), \mathbf{j} = (0, 1, \dots, 0), \dots, \mathbf{z} = (0, 0, \dots, 1) \quad (3)$$

Basis vectors have two important criterion:

- **Linear Independence:** No vector in the set can be written as a combination of the others. When presented with a set of independent vectors  $S = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k)$ , multiplied by some scalar coefficients  $c_1, c_2, \dots, c_n \in R$  the equation,

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + \dots + c_k\mathbf{v}_k = 0, \quad (4)$$

has only one solution, namely that of all  $c_n$  coefficients must be equal to 0. If a non zero solution exists for the coefficients,  $c_n$  it implies that one can write one vector  $\mathbf{v}_n$  as a linear combination of the others, meaning it is dependent on the remaining  $\mathbf{v}_k$  vectors. This would in turn mean that the set of vectors is not linearly independent and hence not a set of base vectors (the n vectors in the set cannot span the entire n dimensional space!).

- **Spanning the Space:** Using the vectors in the basis set, you can build (or “span”) every vector in the entire space by adding them together with appropriate coefficients.

For example, if we consider again the basis vector set  $S = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k)$ , we can say it spans a space  $\mathbf{W}$  if every vector in  $\mathbf{W}$  can be written as a linear combination of the vectors in S, so

$$\mathbf{W} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \dots + a_k\mathbf{v}_k. \quad (5)$$

where  $a_1, a_2, \dots, a_k \in R$  are cofficients which scale vectors in set S.

### 2.3.2 Dot Product

A fundamental operation in vector algebra, the dot product takes two vectors and returns a single number that shows how much one vector “points” in the direction of the other. For example, if you have two vectors,

$$\mathbf{v} = [x_1 \ x_2 \ \cdots \ x_n] \quad \text{and} \quad \mathbf{w} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad (6)$$

their dot product is calculated by multiplying each pair of corresponding components and then summing those products:

$$\mathbf{v} \cdot \mathbf{w} = x_1y_1 + x_2y_2 + \cdots + x_ny_n. \quad (7)$$

This operation is especially useful because it helps us determine the similarity between two vectors, which is an idea that also carries over to functions and signals.

### 2.3.3 Orthogonal Basis

In a vector space, we often work with a set of basis vectors—vectors that can be combined to create any vector in the space. Two vectors are called *orthogonal* if they are at right angles to each other, which means their dot product is zero:

$$\mathbf{v}_i \cdot \mathbf{v}_j = 0 \quad \text{when } i \neq j. \quad (8)$$

Using orthogonal vectors makes many calculations simpler, because each vector contributes independently to any other vector when you express it as a combination of these basis vectors. If we go one step further and ensure that each vector has a length (or magnitude) of 1, the set is called *orthonormal*.

### 2.3.4 Basis Functions

The idea of a basis isn't limited to vectors—it applies to functions as well. Many complex functions can be expressed as a combination of simpler functions called basis functions. For example, suppose you have a function  $x(t)$  that you want to analyze; you can write it as:

$$x(t) = \sum_i a_i \phi_i(t), \quad (9)$$

where each  $\phi_i(t)$  is a basis function and  $a_i$  is a coefficient that tells you how much  $\phi_i(t)$  contributes to  $x(t)$ .

A common example of this is the Taylor series expansion, where a function is expressed as an infinite sum of powers of  $t$ :

$$f(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + \dots = \sum_{n=0}^{\infty} a_n t^n. \quad (10)$$

In this case, the basis functions are simply  $\phi_n(t) = t^n$ .

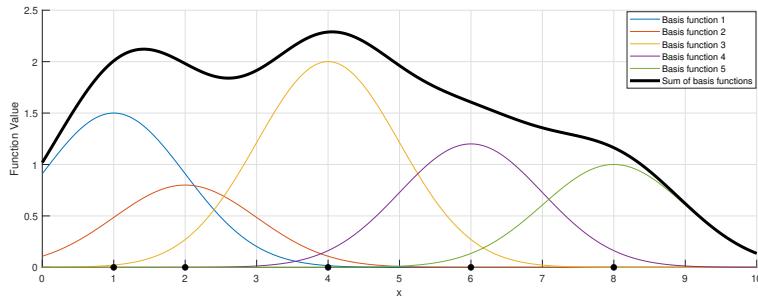


Figure 3: Illustration of how a function (black line) could be decomposed into a sum of 5 basis functions (radial basis) given that the coefficients  $a_1, a_2, \dots, a_5$  are appropriate

### 2.3.5 Hilbert Space: Extending Vector Projections to Functions

Previously, we saw how a vector in three-dimensional space can be “projected” onto orthogonal axes ( $x, y, z$ ). In that finite-dimensional setting, the dot product tells us how much of the vector lies along each axis. This idea naturally extends to *functions* when we move into an infinite-dimensional setting called a **Hilbert space**.

#### From 3D Axes to Function Axes:

In three dimensions, each axis is perpendicular (orthogonal) to the others, so the projection of one axis onto another is zero. Similarly, in a Hilbert space for functions, we think of each “axis” as being one of our orthogonal basis functions. For a periodic signal, these basis functions are often chosen as exponentials  $e^{jk\omega_0 t}$  (or equivalently, sines and cosines).

$$\text{3D Projection: } A_x = \vec{A} \cdot \hat{x} \implies \text{Function Projection: } c_k = \frac{1}{T} \int_0^T x(t) e^{-jk\omega_0 t} dt.$$

In this way, switching from vectors to functions is equivalent to replacing finite sums and dot products with integrals over time. The integer index  $k$  in  $e^{jk\omega_0 t}$  corresponds to each orthogonal “direction” (or harmonic) in the signal’s space. A simple case is shown in the figure below

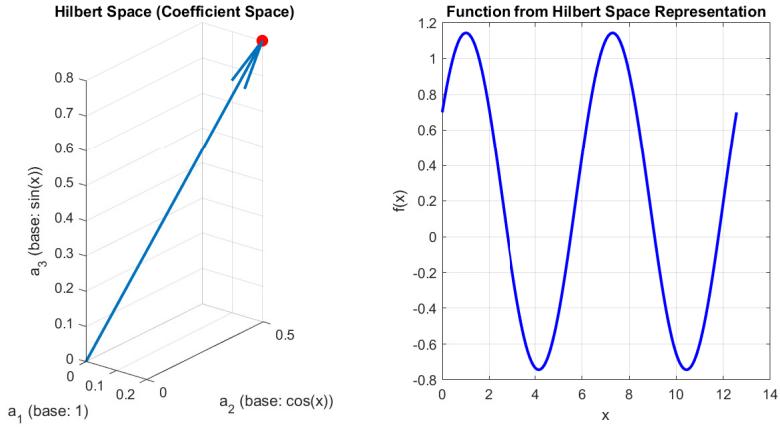


Figure 4: Visualization of Hilbert Space Representation. The left panel displays a 3D vector in a finite-dimensional Hilbert space, with coordinates  $(a_1 = 0.2, a_2 = 0.5, a_3 = 0.8)$  corresponding to the coefficients in the basis  $\{1, \cos(x), \sin(x)\}$ . The right panel shows the function  $f(x) = a_1 + a_2 \cos(x) + a_3 \sin(x)$  constructed as a linear combination of these basis functions over the interval  $[0, 4\pi]$ .

### 2.3.6 Orthogonal/Orthonormal Basis Functions

A set of functions  $\{\phi_i(t)\}$  is said to be *orthogonal* over an interval  $[t_1, t_2]$  if the integral of the product of any two different functions is zero:

$$\int_{t_1}^{t_2} \phi_i(t) \phi_j(t) dt = 0 \quad \text{for } i \neq j. \quad (11)$$

If each function is also normalized—that is, the integral of its square is equal to 1:

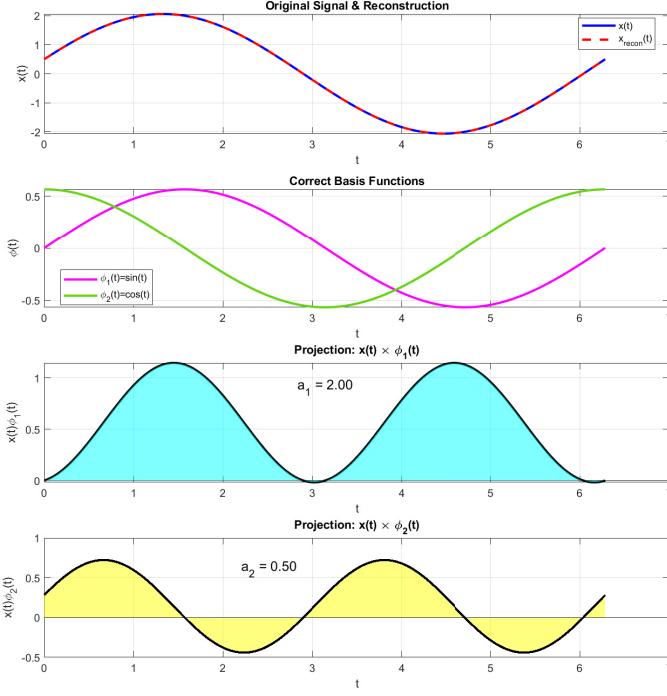
$$\int_{t_1}^{t_2} \phi_i(t)^2 dt = 1, \quad (12)$$

then the set is called *orthonormal*. This is similar to having vectors of length 1, which simplifies many calculations.

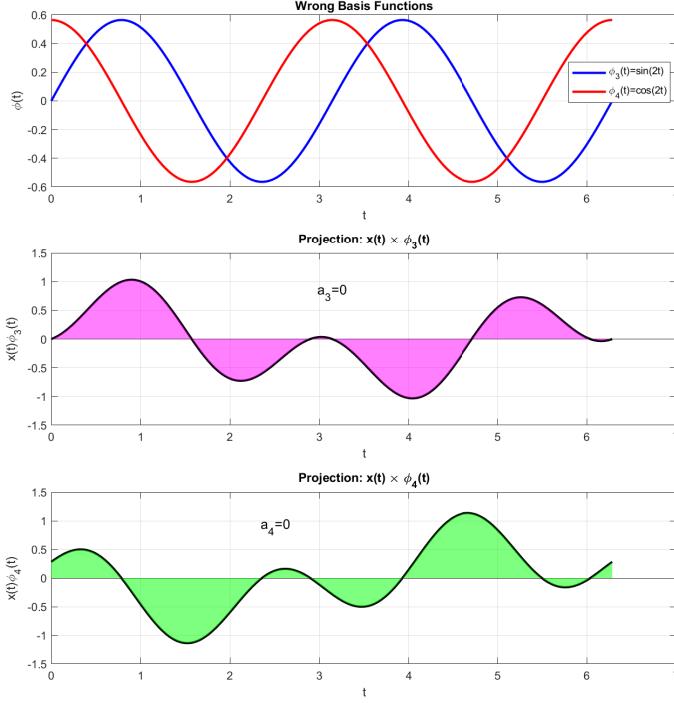
When you represent a function  $x(t)$  as a sum of these basis functions, you determine how much each basis function contributes by “projecting”  $x(t)$  onto each  $\phi_k(t)$ . This is done by calculating:

$$a_k = \frac{1}{\lambda_k} \int_{t_1}^{t_2} \phi_k^*(t) x(t) dt, \quad (13)$$

where  $\lambda_k$  is the value of the integral  $\int_{t_1}^{t_2} \phi_k(t)^2 dt$  (which is 1 for an orthonormal set). This projection gives the coefficient  $a_k$ , telling us the “weight” of each basis function in the overall signal. See Fig. 5 for an illustration of the orthogonal basis function decomposition.



(a) Correct basis: Using  $\sin(t)$  and  $\cos(t)$  to reconstruct  $x(t) = 2\sin(t) + 0.5\cos(t)$  yields significant projection coefficients.



(b) Wrong basis: Projections using  $\sin(2t)$  and  $\cos(2t)$  yield near-zero coefficients, and can be seen that the positive and negative areas cancel each other out

Figure 5: Illustration of orthogonal signal properties. Figure (a) shows the original signal  $x(t) = 2\sin(x) + 0.5\cos(x)$ , and how multiplying function  $x(t)$  with basis functions  $\sin(x)$ ,  $\cos(x)$  and integrating the area of the multiplied signals (cyan and yellow) results in the coefficients of the signals (2 and 0.5 respectively). On the other hand, figure (b) shows that when we use different basis which  $x(t)$  is not composed of, the total area (integration as purple, green) results to 0 (cancels out completely).

## 2.4 Fourier Series

In earlier sections, we introduced the idea of representing vectors and functions with respect to orthogonal (or orthonormal) bases. Recall that each component (or coefficient) in such an expansion is found by projecting the original vector or function onto each of the basis elements.

Now consider a periodic signal  $x(t)$  that repeats every period  $T$ :

$$x(t) = x(t + T). \quad (14)$$

The corresponding fundamental frequency is  $\omega_0 = \frac{2\pi}{T}$ . A powerful way to express  $x(t)$  is to expand it as a sum of orthogonal exponentials:

$$x(t) = \sum_{k=-\infty}^{\infty} c_k e^{jk\omega_0 t}. \quad (15)$$

Here, each  $e^{jk\omega_0 t}$  serves as a *basis function* for the space of all signals with period  $T$ . We call this set of exponentials *orthogonal* because

$$\int_0^T e^{jm\omega_0 t} e^{-jn\omega_0 t} dt = 0 \quad \text{if } m \neq n, \quad (16)$$

mirroring the dot-product (inner-product) property of orthogonal vectors in earlier chapters.

Furthermore, by Euler's formula, each complex exponential can be written as

$$e^{jk\omega_0 t} = \cos(k\omega_0 t) + j \sin(k\omega_0 t). \quad (17)$$

Because of this, we can express  $x(t)$  equivalently in terms of sines and cosines:

$$x(t) = a_0 + \sum_{k=1}^{\infty} [a_k \cos(k\omega_0 t) + b_k \sin(k\omega_0 t)]. \quad (18)$$

These sines and cosines also form an orthogonal set over the interval 0 to  $T$ .

### Relating Coefficients to Projections:

To find each coefficient  $c_k$ ,  $a_k$ , or  $b_k$ , we use the concept of projecting  $x(t)$  onto the corresponding basis function. As described earlier for vectors, you take a “dot product” with the basis vector to isolate that component. In the *function space*, this dot product becomes an integral. Hence, in Fourier analysis :

$$c_k = \frac{1}{T} \int_0^T x(t) e^{-jk\omega_0 t} dt. \quad (19)$$

this integral is the continuous analog of the dot product and directly follows the logic of orthogonal vectors: the only “overlap” each harmonic sees is that which matches its own frequency. This parallels what we did in finite-dimensional vector spaces:

$$A_x = \vec{A} \cdot \hat{x}. \quad (20)$$

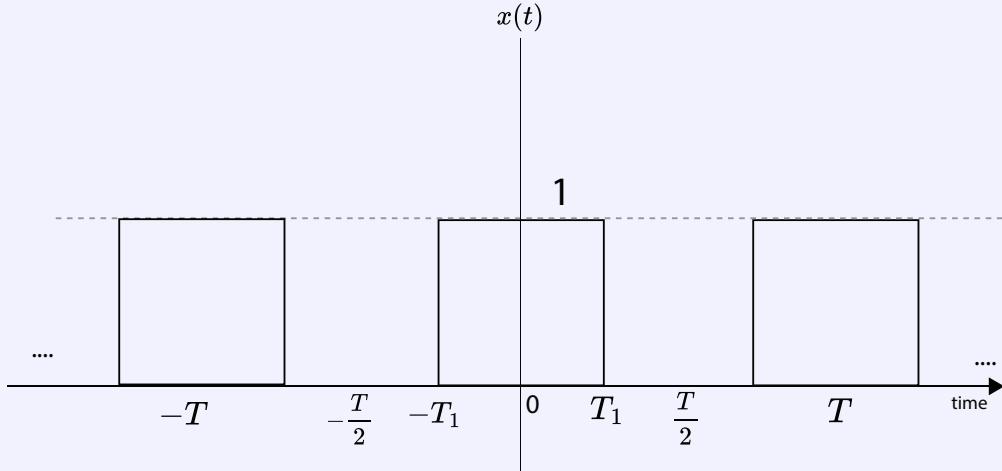
But instead of multiplying components and summing, we multiply functions and integrate over one period. The results give us the individual “weights” of each exponential—or each sine and cosine pair—in the overall signal.

### Interpreting $j$ in Signal Analysis:

While  $j$  is often labeled as the imaginary unit, it's helpful to think of it as indicating a  $90^\circ$  phase shift in the complex plane. Thus,  $j^2 = -1$  represents a  $180^\circ$  rotation, and  $j^4 = 1$  a full  $360^\circ$  rotation back to the real axis. This viewpoint makes it more intuitive to see how sine and cosine functions appear from  $e^{jk\omega_0 t}$ .

### Exercise 1: Computing the Fourier coefficients of a square wave

Suppose we have a periodic square wave such as:



where  $T$  is the period,  $T_1$  is the on-duration, and the square wave has a amplitude of 1. Please compute the Fourier series coefficients and try to draw the frequency representation of the square wave for the first 4 frequency components

#### Solution:

We start by using Eq. (40), to find an analytical form for the coefficients

1. We define the coefficient

$$c_k = \frac{1}{T} \int_{-T/2}^{T/2} x(t) e^{-jk\omega_0 t} dt.$$

where  $\omega_0 = 2\pi \frac{1}{T}$

2. The function  $x(t)$  is nonzero only on  $[-T_1, T_1]$ , so we change the integral limits to

$$c_k = \frac{1}{T} \int_{-T_1}^{T_1} x(t) e^{-jk\omega_0 t} dt.$$

3. Integrate  $e^{-jk\omega_0 t}$  with respect to  $t$ :

$$\int e^{-jk\omega_0 t} dt = \frac{e^{-jk\omega_0 t}}{-jk\omega_0}.$$

4. Evaluate this at  $t = T_1$  and  $t = -T_1$ :

$$c_k = \frac{1}{T} \frac{e^{-jk\omega_0 t}}{-jk\omega_0} \Big|_{t=-T_1}^{t=T_1} = \frac{1}{T} \frac{e^{-jk\omega_0 T_1} - e^{jk\omega_0 T_1}}{-jk\omega_0}.$$

5. Factor out constants and use the identity  $\frac{e^{j\alpha} - e^{-j\alpha}}{2j} = \sin(\alpha)$  and note that  $(\omega_0 = 2\pi f = 2\pi \frac{1}{T})$ . After simplifying, you get

$$c_k = \frac{2}{k \omega_0 T} \cdot \frac{e^{jk\omega_0 T_1} - e^{-jk\omega_0 T_1}}{2j} = \frac{\sin(k\omega_0 T_1)}{k \pi}.$$

6. **Result:** Thus, the final expression is

$$c_k = \frac{\sin(k\omega_0 T_1)}{k \pi}.$$

Now by evaluating the coefficients over the spectrum ( $k = \pm 0, \pm 1, \pm 2, \dots, \pm N$ ), we get the following frequency domain representation of the square wave

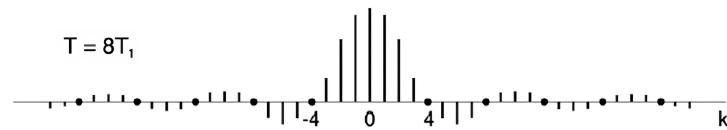


Figure 6: Frequency domain representation of the square wave, assuming  $T = 8T_1$

*If we look closer at the resulting coefficients  $c_k$ , they have the form of a so-called sinc function  $\text{sinc}(x) = \frac{\sin(x)}{x}$ , which will be used often in the course*

The degree to which an oscillating wave with a frequency  $\omega$  is represented in the spectrum of a given function  $f(t)$  can be calculated by finding the area under the graph of the multiplication.

In the Fourier series equations, this is in fact the area of overlap between a signal and a candidate sinusoidal function (this is why the term  $\omega$  is present inside cosine, Eq. ??). Each frequency (multiples of  $2\pi$ ) has its own corresponding integral result (area calculation), hence  $F(\omega)$ . The actual coefficient, so weight of each sinusoid is thus the ratio between the computed area and the effective area of the stand-alone sinusoid (within the same interval).

Figure 7 showcases the area of correlation described by the Fourier integral. As a follow-up, Figures 8, 9 and 10 indicate the same process, now at separate frequencies  $\omega$ .

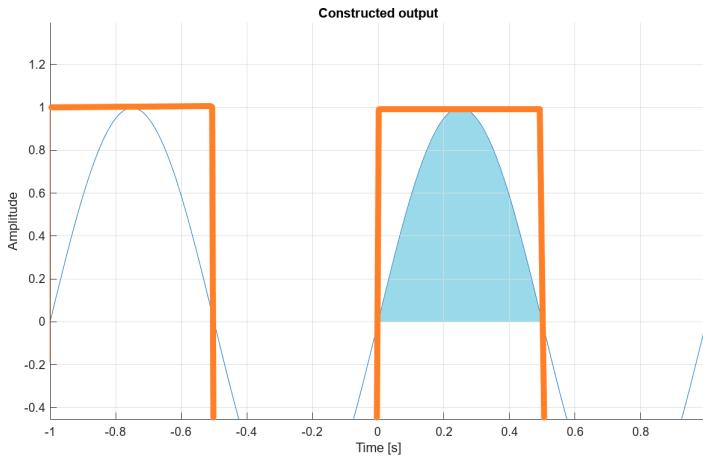


Figure 7: Square wave  $\times$  1 Hz sinusoid, with shaded net area—what does this nonzero area tell us about the 1 Hz component?

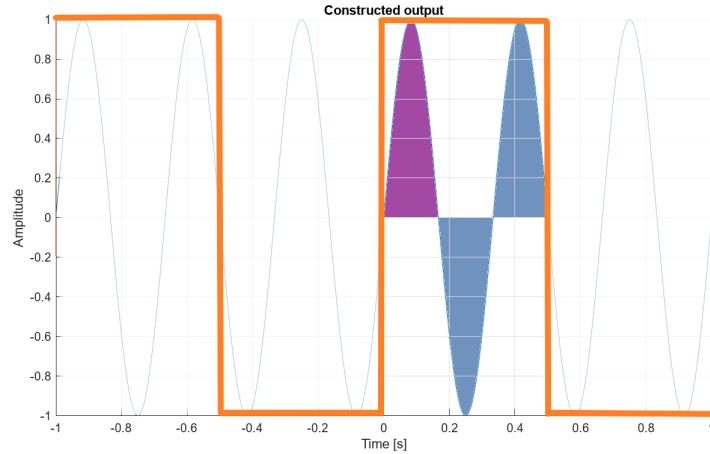


Figure 8: What frequency does this term have? Moreover, what does the net area (shaded) tell us?

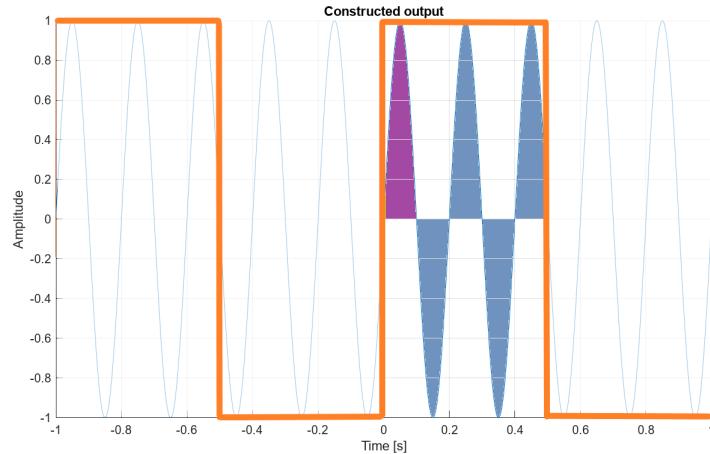


Figure 9: What about this one?

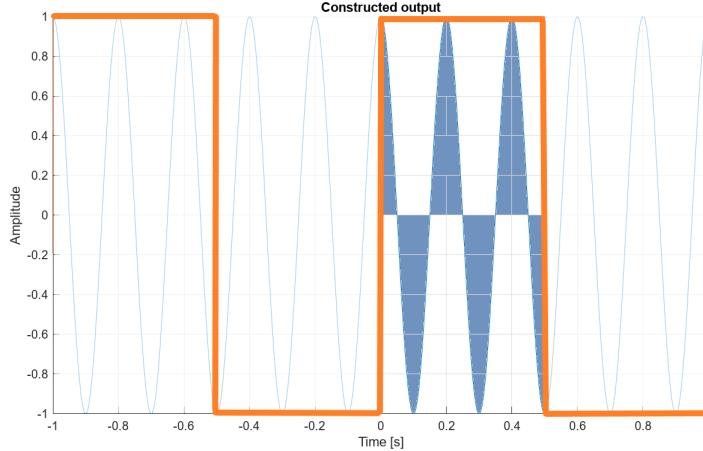


Figure 10: Why are all the areas computed cancelling each other?

#### 2.4.1 Quadrature Form to Obtain Fourier Series Coefficients

Periodic signals—such as rectangular, square, or sawtooth waves—often exhibit symmetries that allow them to be represented using a single set of basis functions (sine or cosine), with the only difference being a  $90^\circ$  phase shift. The Fourier series reconstructs a periodic signal by combining its DC component and harmonic contributions as follows:

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[ a_n \cos\left(\frac{n\pi t}{L}\right) + b_n \sin\left(\frac{n\pi t}{L}\right) \right] \quad (21)$$

Here, the parameter  $L$  is defined as **half** the period of the signal ( $L = \frac{T}{2}$ ). The coefficients are determined by quadrature integrals over the symmetric interval  $[-L, L]$ .

The DC component can be found as:

$$a_0 = \frac{1}{T} \int_{-L}^L f(t) dt \quad (22)$$

Note that for the DC component we divide by total period  $T$  instead of half of the period  $L$ . And the cosine and sine coefficients are computed as:

$$a_n = \frac{1}{L} \int_{-L}^L f(t) \cos\left(\frac{n\pi t}{L}\right) dt \quad (23)$$

$$b_n = \frac{1}{L} \int_{-L}^L f(t) \sin\left(\frac{n\pi t}{L}\right) dt \quad (24)$$

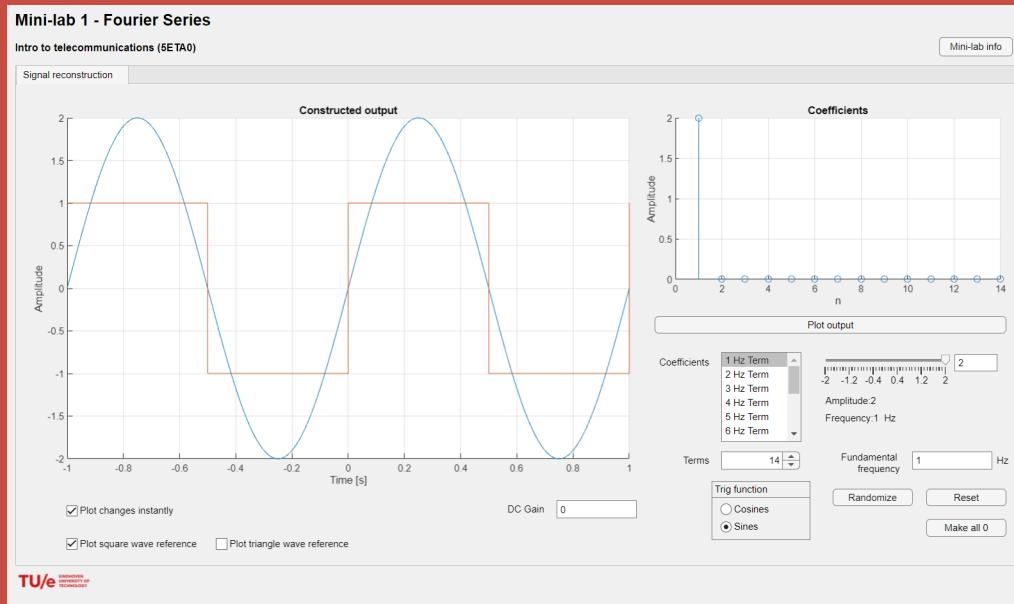
## Mini-lab exercise 2.1 - Fourier coefficients

To get started with the minilab, please see section 1.3 of the reader for a starting point.

Below you can find the set-up of the Fourier series Mini-lab. As can be seen, you have the option of manually inputting the  $a_k$  and  $b_k$  coefficients. This then allows you to create various shapes in signal plotter, and thus you can reconstruct practically any signal within the platform. Furthermore, you can choose whether to plot with sin or cosine terms, and you may also plot a reference signal.

To become familiar with the app, manually calculate the first five Fourier coefficients (sine), for the square wave signal in the figure.

Insert these coefficients into the MATLAB app, and reason whether you get the expected result.



For the answers obtained in the previous exercise, try plotting them in the MATLAB app, now selecting the option of plotting using cosine functions instead. What changes are present in the reconstructed signal? If the result is not as expected, what should you change with regard to your coefficients? What properties don't the cos and sin functions share?

In the case that you cannot provide an answer, try re-calculating the coefficients and input your results in the app (of course, you are welcome to use an online tool for re-doing the computations).

## 2.5 Fourier Transform

For functions that are not periodic, or when we consider the limit as the period  $T \rightarrow \infty$ , the Fourier series representation transforms into what is known as the **Fourier transform**. The Fourier transform of a function  $f(t)$  is defined as

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt, \quad (25)$$

where  $F(\omega)$  is a complex-valued function that represents the amplitude and phase of the frequency component  $\omega$ .

The integral in Equation 25 essentially decomposes the function  $f(t)$  into an infinite number of sinusoidal components. Unlike the Fourier series, which sums discrete frequency components, the Fourier transform integrates over a continuous range of frequencies. The figure below shows a basic example of the Fourier transform applied to a non-periodic square wave

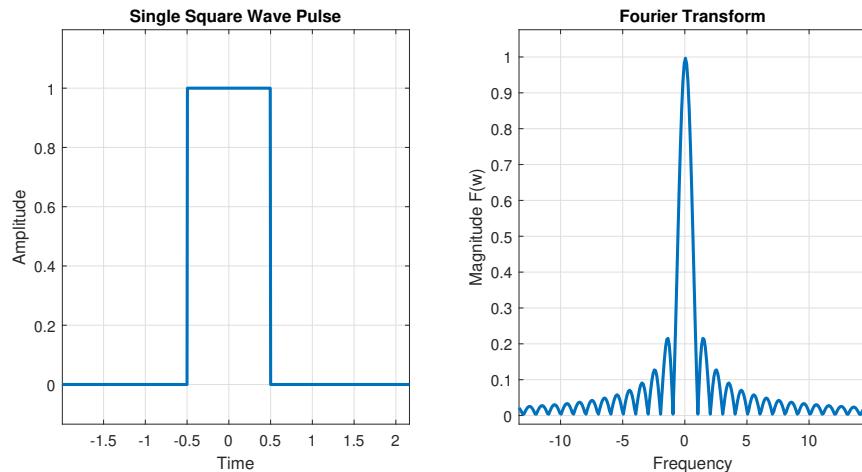


Figure 11: A single square wave pulse and its Fourier transform (a sinc). Note that the square wave does not repeat in time, hence its non-periodic signal.

### Exercise 2: Computing the Fourier Transform of a Square Pulse

Consider the square pulse defined by

$$x(t) = \begin{cases} 1, & |t| \leq 0.5, \\ 0, & |t| > 0.5. \end{cases}$$

Compute the Fourier transform  $X(\omega)$  of  $x(t)$ , i.e.,

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt.$$

Also, sketch the magnitude of  $X(\omega)$  for a few values of  $\omega$ .

**Solution:**

1. **Restricting the Integration Limits:** Since  $x(t) = 0$  for  $|t| > 0.5$ , the integral reduces to:

$$X(\omega) = \int_{-0.5}^{0.5} e^{-j\omega t} dt.$$

2. **Integrate:** The integral of the complex exponential is:

$$\int e^{-j\omega t} dt = \frac{e^{-j\omega t}}{-j\omega}.$$

Evaluating from  $t = -0.5$  to  $t = 0.5$  gives:

$$X(\omega) = \left. \frac{e^{-j\omega t}}{-j\omega} \right|_{t=-0.5}^{0.5} = \frac{e^{-j\omega(0.5)} - e^{j\omega(0.5)}}{-j\omega}.$$

3. **Simplify Using Euler's Formula:** Recall the identity

$$e^{-j\theta} - e^{j\theta} = -2j \sin(\theta).$$

Setting  $\theta = \omega/2$  leads to:

$$e^{-j\omega(0.5)} - e^{j\omega(0.5)} = -2j \sin\left(\frac{\omega}{2}\right).$$

Therefore,

$$X(\omega) = \frac{-2j \sin\left(\frac{\omega}{2}\right)}{-j\omega} = \frac{2 \sin\left(\frac{\omega}{2}\right)}{\omega}.$$

4. **Final Expression:** The Fourier transform of the square pulse is:

$$X(\omega) = \frac{2 \sin\left(\frac{\omega}{2}\right)}{\omega}.$$

This expression is often written in terms of the sinc function (defined here as  $\text{sinc}(x) = \frac{\sin(x)}{x}$ ):

$$X(\omega) = \text{sinc}\left(\frac{\omega}{2}\right).$$

**Frequency Domain Representation:**

The magnitude of  $X(\omega)$  is given by:

$$|X(\omega)| = \left| \frac{2 \sin\left(\frac{\omega}{2}\right)}{\omega} \right|.$$

This function has a main lobe centered at  $\omega = 0$  with successive zero crossings at  $\omega = \pm 2\pi, \pm 4\pi, \dots$ . The shape is characteristic of a sinc function, with side lobes that decay as  $|\omega|$  increases.

**Remark:** The Fourier series is used to represent periodic signals as a sum of sinusoids at discrete frequencies (i.e., the spectrum is discrete). In contrast, the Fourier transform applies to non-periodic signals and yields a continuous spectrum. This can be viewed by comparing the solutions of this exercise and exercise 1

### 2.5.1 Properties of the Fourier transform

- **Time Shift:**

- Time shifting in the time domain corresponds to phase shifting in the frequency domain.
- If  $x(t) \xrightarrow{\mathcal{F}} X(f)$ , then  $x(t - t_0) \xrightarrow{\mathcal{F}} e^{-j2\pi f t_0} X(f)$ .

- **Frequency Shift:**

- Multiplication with a time dependent linear phase shift results in a frequency shift in the frequency domain.
- If  $x(t) \xrightarrow{\mathcal{F}} X(f)$ , then  $e^{j2\pi f_0 t} x(t) \xrightarrow{\mathcal{F}} X(f - f_0)$ .

- **Time Scaling:**

- Compression/expansion in the time domain results in expansion/compression in the frequency domain.
- If  $x(t) \xrightarrow{\mathcal{F}} X(f)$ , then  $x(at) \xrightarrow{\mathcal{F}} \frac{1}{|a|} X\left(\frac{f}{a}\right)$ .

- **Frequency Scaling:**

- Compression/expansion in the frequency domain results in expansion/compression in the time domain.
- If  $X(f) \xrightarrow{\mathcal{F}^{-1}} x(t)$ , then  $|a|X(af) \xrightarrow{\mathcal{F}^{-1}} x\left(\frac{t}{a}\right)$ .

- **Time Reversal:**

- Time reversal in the time domain corresponds to complex conjugation in the frequency domain.
- If  $x(t) \xrightarrow{\mathcal{F}} X(f)$ , then  $x(-t) \xrightarrow{\mathcal{F}} X^*(-f)$ .

- **Frequency Reversal:**

- Reversing the spectrum in the frequency domain corresponds to time reversal in the time domain.
- If  $X(f) \xrightarrow{\mathcal{F}^{-1}} x(t)$ , then  $X(-f) \xrightarrow{\mathcal{F}^{-1}} x(-t)$ .

- **Convolution Theorem:**

- Convolution in the time domain corresponds to multiplication in the frequency domain.
- If  $x(t) \xrightarrow{\mathcal{F}} X(f)$  and  $y(t) \xrightarrow{\mathcal{F}} Y(f)$ , then  $x(t) * y(t) \xrightarrow{\mathcal{F}} X(f) \cdot Y(f)$ .

- **Multiplication in Time:**

- Multiplying by a function in the time domain corresponds to convolution in the frequency domain.
- If  $x(t) \xrightarrow{\mathcal{F}} X(f)$  and  $h(t) \xrightarrow{\mathcal{F}} H(f)$ , then  $x(t) \cdot h(t) \xrightarrow{\mathcal{F}} X(f) * H(f)$ .

- **Differentiation in Time:**

- Differentiation in the time domain corresponds to multiplication by  $j2\pi f$  in the frequency domain.
- If  $x(t) \xrightarrow{\mathcal{F}} X(f)$ , then  $\frac{d}{dt}x(t) \xrightarrow{\mathcal{F}} (j2\pi f)X(f)$ .

- **Integration in Time:**

- Integration in the time domain corresponds to division by  $j2\pi f$  in the frequency domain (with a caveat for the DC component).
- If  $x(t) \xrightarrow{\mathcal{F}} X(f)$ , then  $\int_{-\infty}^t x(\tau)d\tau \xrightarrow{\mathcal{F}} \frac{X(f)}{j2\pi f} + \pi X(0)\delta(f)$ .

## 3 Sampling theory

### 3.1 Learning objectives

Students completing this chapter should have learned:

1. Understand the concept of band limited vs time limited signals.
2. Can sketch the spectrum and time evolution of an ideal sampled signal.
3. Can calculate and motivate the required sampling period (frequency) for a given signal with a defined bandwidth.
4. Can evaluate the chance of aliasing and understand the impact aliasing can have on signal reconstruction.
5. Understand why a *Sinc* interpolation function allows to perfectly reconstruct a sample signal from its samples.
6. Understand the limitations of ideal sampling concept and why it cannot be used in practice.
7. Be able to calculate the maximum number of independent pieces of information needed for describing a waveform of duration  $T_0$  and bandwidth  $B$ .

### 3.2 Motivation

The first step towards going into the digital domain is the question of how do we convert an analog signal which is, for example, a voltage varying through time (in our example, the signal is a voltage coming from a microphone), into a stream of binary numbers that represent the analog signal. The answer to this question is that we have to convert the analog signal to a digital representation by sampling it at specific intervals. In this section, you will learn about the ideal sampling method, analyze it in the time domain and frequency domain, and derive a mathematical description for the ideal sampling method.

Furthermore, when you have a sampled signal, it is crucial to understand how you can recover the original analog signal from the samples. In this chapter we show that if we adhere to some basic guidelines of the sampling theorem, we are guaranteed to be able to reconstruct the original analog signal from its samples during the signal recovery phase without any loss of information or resolution!

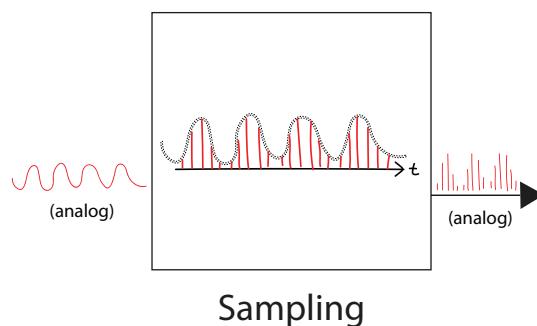
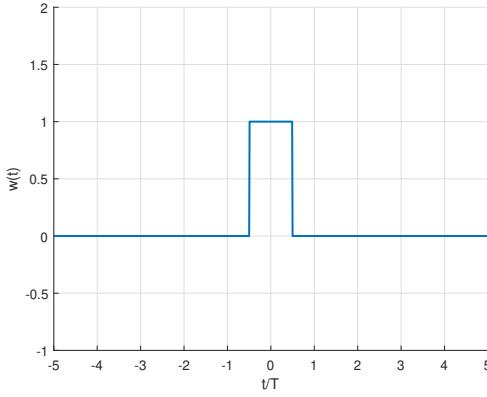
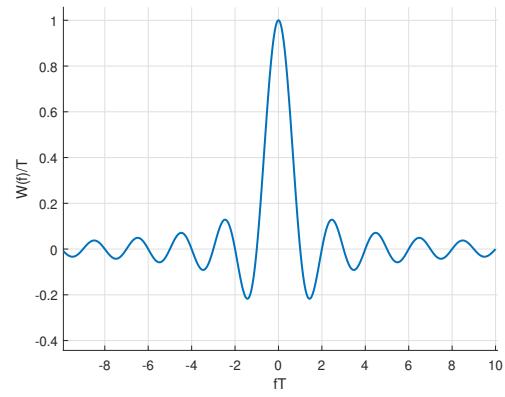


Figure 12: Sampling theory - Topic map location



(a) The rectangular pulse in time domain



(b) The frequency spectrum of the rectangular pulse

Figure 13: Plot of a rectangular pulse and its Fourier transform, for  $T = 1\text{s}$

### 3.3 Band-limited and time-limited signals

A **band-limited** signal has non-zero frequency components only for a limited part of the frequency spectrum, and the rest of the frequency spectrum is zero. As a result the same signal, when observed in the time domain, will have an unlimited time span. For example, a cosine function has a single frequency component but exists for infinite time.

On the other hand, a **time-limited** signal has non-zero amplitude only for a limited part of the time (meaning it exists for only some time, but not forever). Consequently, it will have an infinite (or unlimited) spectral representation. For example a step function does not exist in negative time, hence will need an infinite amount of frequencies to construct. Hence, we can say,

**A waveform is either time-limited or band-limited; not both!**

*Although the signal may be strictly speaking band unlimited (there are non-zero amplitude in the spectrum for all possible values of  $f$ ), it may be bandlimited for all practical purposes in the sense that the amplitude spectrum has a negligible level above a certain frequency since the amplitude of those frequency components is very low.*

#### 3.3.1 Illustration of band-limited and time-limited signals

Suppose in the time domain we have a rectangular pulse, which exists from a time interval of  $-T/2 < t < T/2$  (hence time-limited) and can be denoted as

$$w(t) = \prod\left(\frac{t}{T}\right) = \begin{cases} 1 & \text{for } |t| < \frac{T}{2} \\ 0 & \text{otherwise} \end{cases} \quad (26)$$

This pulse is visualized in Fig. 13a. Given the pulse in time domain, we want to find the frequency components that make it up, hence we can take its Fourier transform of  $w(t)$  which we denote as  $W(f)$ .

Similar to exercise 2 from the previous chapter,

$$W(f) = FT\{w(t)\} = \int_{-\infty}^{\infty} w(t)e^{-j\omega t} dt = \int_{-\frac{T}{2}}^{\frac{T}{2}} 1 e^{-j\omega t} dt \quad (27)$$

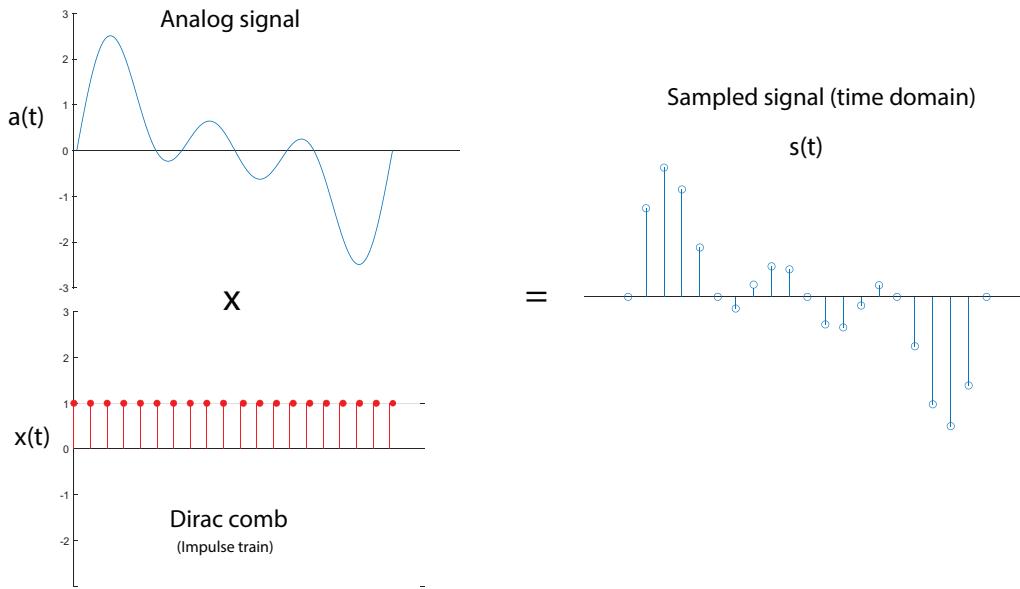


Figure 14: An analog signal being multiplied by a Dirac comb (impulse train), and the resulting sampled signal in time domain

Evaluating this integral and using Euler's formula we end up with

$$W(f) = T \frac{\sin(\omega T/2)}{\omega T/2} = Tsinc(fT) = TSa(\pi fT) \quad (28)$$

Note that we define

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}, \quad Sa(x) = \frac{\sin(x)}{x} = \text{sinc}\left(\frac{x}{\pi}\right) \quad (29)$$

If we plot Eq. (28) we get Fig. 13b, we can observe that a sinc function exists from  $-\infty$  to  $\infty$ , meaning there are an unlimited amount of frequency components.

*however, as stated before for practical real-life purposes the very high frequencies will have amplitudes so small, we can neglect them.*

### 3.4 Impulse sampling & Sampling theorem

Finally, with the knowledge of band-limited and time-limited waveforms, we can develop a **sampling theorem**. Let us start from the beginning and try to build an ideal sampling method and theorem

We start by considering an analog signal (which is a voltage changing over time), and we assume this signal is coming from a microphone (ex. see the microphone on the topic map Fig. 1). Our sampling methodology must fulfill the following requirement:

- We must make sure that the original analog signal can again be reconstructed at the end, from its sampled form, (ex. see the last block on the topic map Fig. 1). This means that we need a sample-rate higher or equal to twice the maximum frequency in the signal. This is called the Nyquist criterium and will be expanded on later.

Ideal sampling is done by multiplying a signal with an "impulse Dirac comb", which results in an impulse sampled waveform. This concept is illustrated in Fig. 14.

Please note that this sampled signal is still in the continuous time domain and NOT discrete time domain, because the sampled signal still has a time interval between samples. The next step in the communication chain is quantization, where this sampled signal gets quantized and encoded into a stream of binary digits, and then it becomes discrete time.

We now focus on analyzing the impulse Dirac comb in the frequency domain, and with the help of the properties of the Fourier transform we will try to derive a mathematical expression for the sampled signal. We can mathematically denote the Dirac comb as

$$x(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT) \quad (30)$$

where  $\delta(t)$  is a unit impulse defined as

$$\delta(t) = \begin{cases} 1 & \text{for } t = 0 \\ 0 & \text{for } t \neq 0 \end{cases} \quad (31)$$

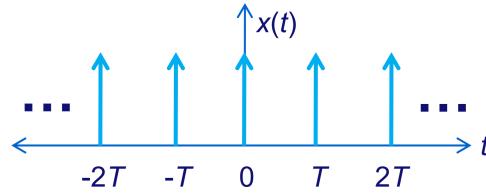


Figure 15: Dirac comb train (impulse train) visualized in continuous time-domain

where  $T$  is the sampling period. Visually, Eq. (30) is illustrated in Fig. 15. Equivalently, we may show the Dirac impulse comb as a sum of complex exponentials (see link [1] for intuitive visualizations)

$$x(t) = \sum_{k=-\infty}^{\infty} c_k e^{j k \omega_0 t} \quad \text{where } \omega_0 = \frac{2\pi}{T} \quad (32)$$

and where  $c_k$  are the complex Fourier series coefficients. To find these coefficients, we may use the Fourier series coefficient equation

$$c_k = \frac{1}{T} \int_{-T/2}^{T/2} x(t) e^{-jk\omega_0 t} dt = \frac{1}{T} \int_{-T/2}^{T/2} \delta(t) e^{-jk\omega_0 t} dt = \frac{1}{T} [e^{-jk\omega_0 t}]_{t=0} = \frac{1}{T} \quad (33)$$

Substituting  $c_k$  in Eq. (32) from the result found in Eq. (33) our Dirac comb as a sum of complex exponentials may be represented as

$$x(t) = \sum_{k=-\infty}^{\infty} \frac{1}{T} e^{j k \omega_0 t} \quad (34)$$

Now the impulse sampled signal can be written as (using Eq. (30))

$$s(t) = a(t)x(t) = \sum_{n=-\infty}^{\infty} a(t)\delta(t - nT) = \sum_{n=-\infty}^{\infty} a(nT)\delta(t - nT) \quad (35)$$

where  $s(t)$  is the sampled output signal and  $a(t)$  is the analog input signal. Now we have an expression for the sampled signal in Fig. 14. However, we can also write the sampled signal  $s(t)$  using Eq. (34) for the Dirac comb instead of Eq. (30), which then yields to

$$s(t) = a(t)x(t) = \frac{1}{T} \sum_{k=-\infty}^{\infty} a(t)e^{-jk\omega_0 t} \quad (36)$$

Now we can take the Fourier Transform of Eq. (36) to finally find the frequency spectrum  $S(\omega)$  of the impulse sampled signal  $s(t)$

$$S(\omega) = \mathcal{F}\{s(t)\} = \mathcal{F}\{a(t)x(t)\} = \mathcal{F}\left\{\frac{1}{T} \sum_{k=-\infty}^{\infty} a(t)e^{-jk\omega_0 t}\right\} \quad (37)$$

We may rewrite Eq. (37) as

$$S(\omega) = \frac{1}{T} \sum_{k=-\infty}^{\infty} \mathcal{F}\{a(t)e^{-jk\omega_0 t}\} \quad (38)$$

To further evaluate the  $\mathcal{F}\{\dots\}$ , we may recall a property of Fourier transform that is given as

$$a(t)e^{-jk\omega_0 t} \xrightarrow{\text{Fourier Transform}} A(\omega - k\omega_0) \quad (39)$$

which means that a multiplication by a complex exponential in the time domain is equivalent to a shift in frequency in the frequency domain. By using this property we may evaluate Eq. (38) further and find that

$$\boxed{S(\omega) = \frac{1}{T} \sum_{k=-\infty}^{\infty} A(\omega - k\omega_0)} \quad (40)$$

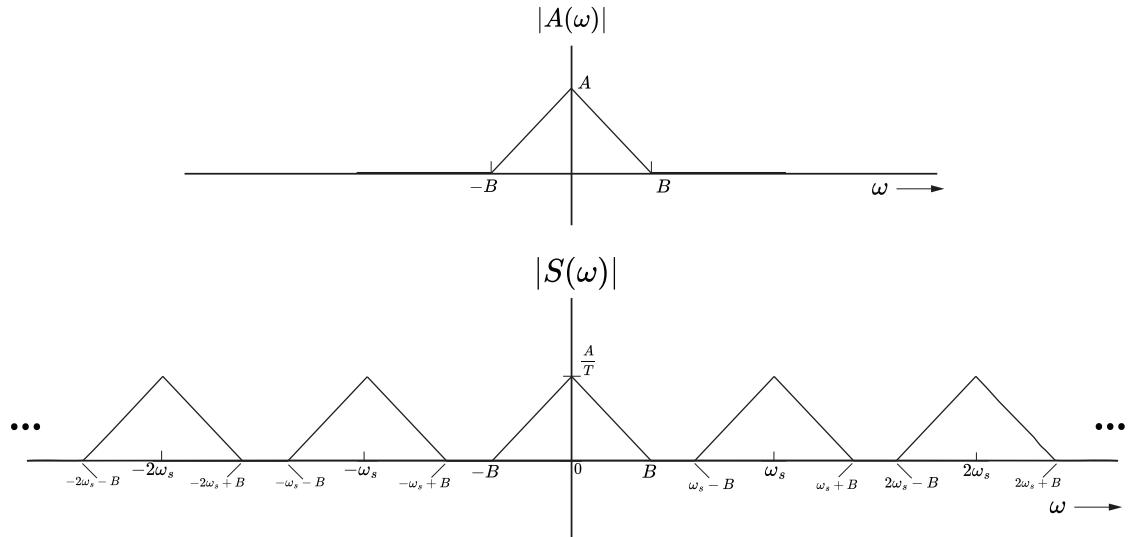


Figure 16: The concept of spectrum repetitions [2, ch.2-7]

Eq. (40) is an interesting finding because it shows that the initial spectrum of the analog signal  $A(f) = \mathcal{F}\{a(t)\}$  is being repeated infinitely over the frequency domain every multiple

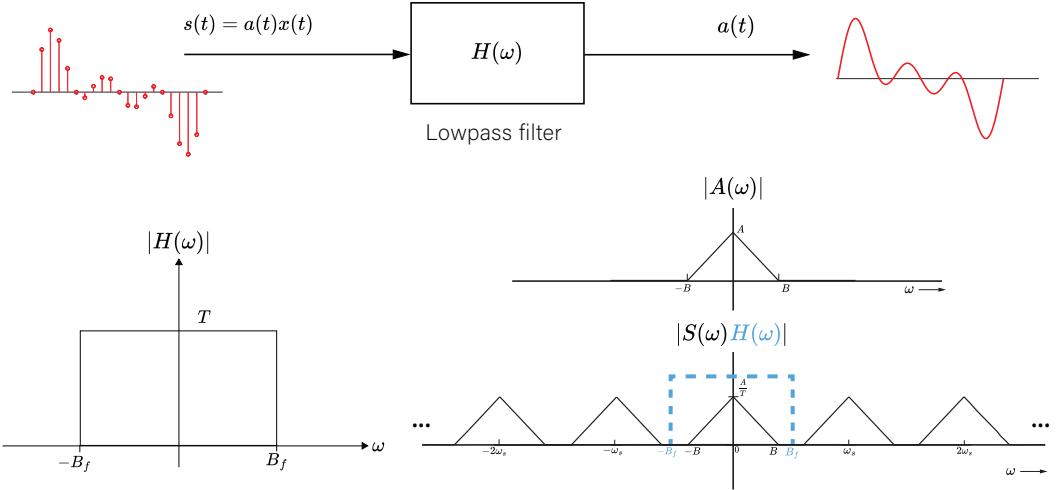


Figure 17: The impuled sampled signal filtered with a ideal square lowpass filter for exact reconstruction

of the sampling frequency  $\omega_0 = 2\pi f_s = 2\pi/T$ , due to the effect of impulse sampling. If we assume  $A(f)$  has a triangular spectrum with bandwidth  $B$ , then we can illustrate the effect of the repeated spectrum in Fig. 16. Moreover, we can observe that the amplitude of the frequency spectrum will be multiplied by  $1/T$ .

Having sampled the original signal (defined as  $s(t)=a(t)x(t)$ ), the subsequent question is: How can the original signal  $a(t)$  be accurately retrieved from its sampled form  $s(t)$ ?

### 3.4.1 Signal reconstruction

If we look at Fig. 16, we see that to be able to get to the original signal  $a(t)$ , we need to only extract one of the spectral copies (triangle). The easiest way would be to pick the copy centered around  $f=0\text{Hz}$ . However, we must make sure that we don't extract other copies of the spectrum or parts of them, as they may cause additional higher frequency components which will then result in the original signal not being reconstructed properly.

To extract the copy around the center we can use a simple filter (a low pass filter). We can also pick any other copy that is not at the center, however, that will require additional signal processing steps which will complicate the reconstruction.

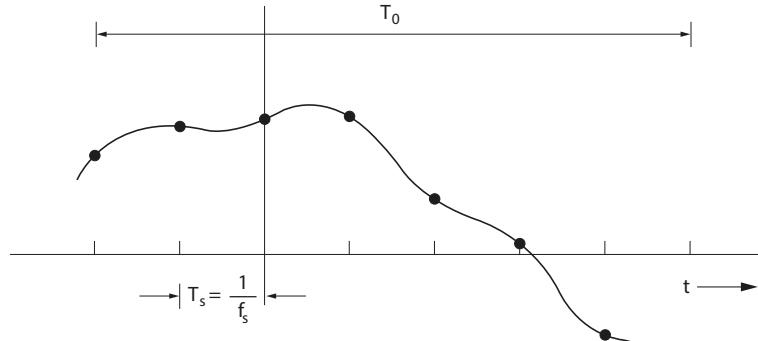
Lets see how the use of an ideal low pass filter allow us to perfectly reconstruct the original signal. An ideal low pass filter would have a rectangular spectrum with bandwidth  $B$ , centered around zero, which extends exactly to only extract the first repetition around zero, and reject all the other repetitions. This concept is illustrated in Fig. 17.

To denote the lowpass filter in the frequency domain we can simply write it as

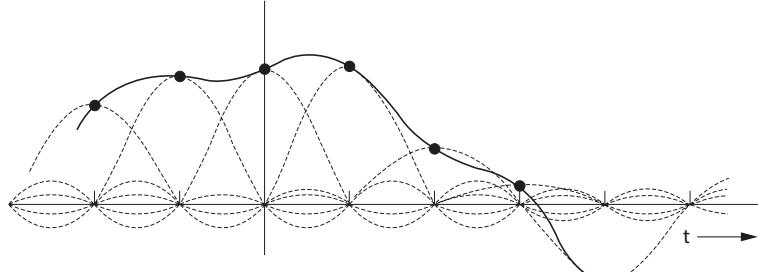
$$H(\omega) = \begin{cases} T & \text{for } -B \leq \omega \leq B \\ 0 & \text{otherwise} \end{cases} \quad (41)$$

where  $T$  is the sampling period (*the reason the amplitude of this low-pass filter is  $T$  is to cancel out the  $1/T$  term caused by impulse sampling as seen on Eq. (40)*).

The impulse response of this filter can be determined by taking the inverse Fourier transform



(a) Waveform and Sample Values



(b) Waveform Reconstructed from Sample Values

Figure 18: Ideal Sampling theorem [2, ch. 2-7]

of (41), and is found to be

$$h(t) = \frac{BT}{\pi} \operatorname{sinc}\left(\frac{Bt}{\pi}\right) \quad \text{where } -\infty < t < \infty \quad (42)$$

Furthermore, the reconstruction of a discrete-time signal into continuous time can be achieved by filtering the sampled waveform  $s(t)$  with the ideal filter  $h(t)$ , which leads to the original signal  $a(t)$

$$a(t) = \sum_{n=-\infty}^{\infty} a_n \frac{\sin\{\pi f_s[t-(n/f_s)]\}}{\pi f_s[t-(n/f_s)]} \quad (43)$$

where  $a_n$  is the  $n$ -th sample of the original sampled signal. Figure 18 shows an example of a signal being sampled (Fig. 18a), and then reconstructed later from the sample points via Eqn. 43 (Fig. 18b). We can see that in the time domain, the interpolation filter is **non-causal**, meaning it is also defined in time for  $t < 0$ . This is clearly not something we can implement in real-time, but can be done if processing is done "off-line". Hence theoretically based on this formula one can reconstruct a signal from its samples perfectly. In practical systems, and those working in real time simpler and causal interpolation functions must be used leading to sub optimal reconstruction. Such solutions are beyond the scope of this course.

### 3.4.2 Aliasing

Another observation one can make from studying Eq. (40) is that if the sampling frequency  $f_s$  is too small, the repeated spectra will overlap.

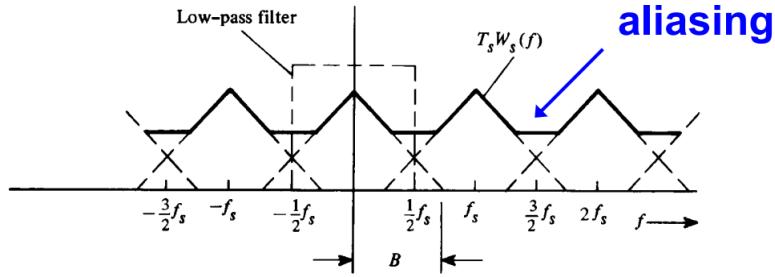


Figure 19: Visualization of the effect of aliasing in frequency domain

The spectra overlapping is visualized in Fig. 19. As can be seen, this overlapping will cause our original spectrum to be deformed, hence causing a loss of information about our original signal and making it impossible to recover the original signal. This effect is known as **aliasing**.

In order to prevent aliasing, by looking at the repeated spectra, we must make sure that the sampling frequency  $\omega_0 = 2\pi f_s$  is at least twice as big as the bandwidth  $B$  of the signal being sampled. This relation is known as **Nyquist's sampling rate criteria**, and it can be formally written as

$$f_s \geq 2B \quad (44)$$

where  $f_s$  is the sampling frequency and  $B$  is the bandwidth (or the highest frequency component) of our signal which is being sampled.

**We must always pick the sampling frequency such that it satisfies Nyquist's sampling rate to avoid aliasing**

*You may have wondered why audio is usually sampled at a sampling rate of 44.1-48 kHz, which is a familiar number in audio systems, the reason is because of Nyquist sampling rate.*

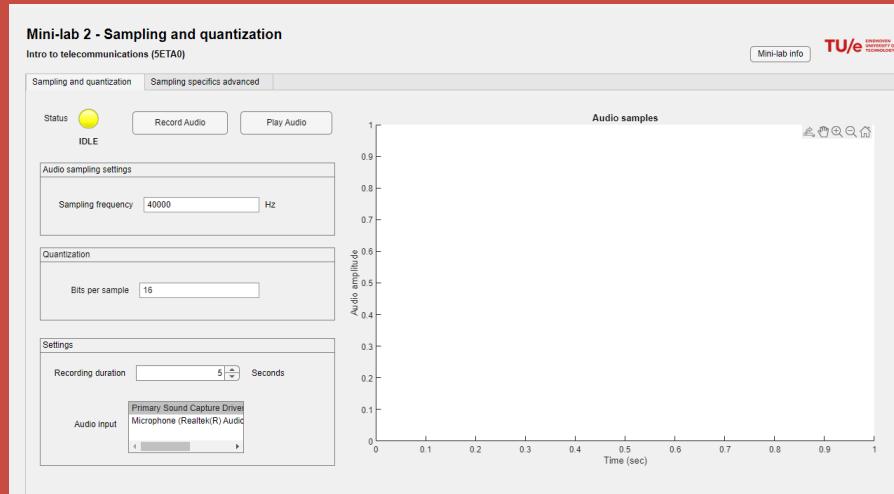
*The human hearing range is between 20 Hz - 20 kHz, hence the hearable spectrum is  $B = 20$  kHz, so in order to not lose any audio information (hence prevent aliasing), we need to sample audio inputs at  $2B \approx 44.1$  kHz*

## Minilab exercise 3.1 - Aliasing demonstration in time domain

To get started with the minilab, please see section 1.3 of the reader for a starting point.

This mini-lab exercise requires you to use Mini-lab 2 (Sampling and Quantization) on MATLAB.

When you open Minilab 2 you will be confronted with the following view:



In the settings section, you may change the microphone recording duration and also which microphone to use from your computer. Furthermore, on the *audio sampling settings* you may change the sampling frequency, and for this exercise, that is of interest. Leave the number of bits per sample to the default configuration of 16 (*bits per sample will be introduced later in the course*)

- 1) Start by recording your voice with the default configuration by pressing the 'record audio button'. After you see the audio waveform on the screen, click on *play audio* to make sure that the minilab is working on your computer
- 2) Now, supposing that you completed the first step and the microphone is functioning with the minilab, change the sampling frequency to 20000 Hz, record and play the audio again, what changes in the audio do you now notice?
- 3) Now, to observe the aliasing effect better, start at 10000 Hz, record and play, and on each iteration, decrease the frequency by 2500 Hz, until you get to 2500 Hz. Using the concept of aliasing and the audio you are hearing as you decrease the sampling frequency, can you explain why the audio is getting distorted?
- 4) What sampling frequency would you choose to prevent the aliasing of microphone-recorded audio? Choose that sampling frequency and increase it to around 70000 Hz on steps of 5000 Hz, do you observe any significant change in the audio as you are increasing the sampling frequency?

### 3.5 Dimensionality theorem

The dimensionality theorem describes the number of **independent** pieces of information, which can describe a real waveform. That number  $N$  is mathematically described as

$$N = 2BT_0 \quad (45)$$

where  $B$  is the bandwidth of the waveform and  $T_0$  is the time span over which the signal is to be described (sampled). Essentially, through the dimensionality theorem, we are connecting the number of data points (pieces of information) to the sampling frequency and the time interval of sampling. It tells us that we cannot get more information by adding more points (oversampling) since then the information in these additional points is no longer independent!

In other words, the information that can be conveyed by a band-limited waveform (with a bandwidth  $B$ ) or a band-limited communication system is proportional to the product of the bandwidth of that signal/system and the time allowed for transmission of the information.

#### Exercise 3: Dimensionality theorem

Consider a signal with a bandwidth of

$$B = 64 \text{ kHz},$$

and suppose we wish to reconstruct this signal over a duration of

$$T = 10 \text{ seconds.}$$

Determine:

1. The number of symbols (independent pieces of information) required for reconstruction
2. The total amount of information in bits conveyed by the signal, assuming each symbol encodes 4 bits of information (*will be discussed more in detail during the later sections*).

**Solution:**

1. **Number of Symbols:**

According to Eq. 42, the number of symbols  $N$  is given by:

$$N = 2BT.$$

Substituting the given values:

$$N = 2 \times (64 \text{ kHz}) \times (10 \text{ s}) = 1280 \text{ kSymbols.}$$

This indicates that 1280 thousand independent symbols are used to represent the signal.

2. **Total Information Conveyed:**

If each symbol carries 4 bits of information, then the total information  $I$  (in kilobits) is:

$$I = 1280 \text{ kSymbols} \times 4 \text{ bits/symbol} = 5120 \text{ kBits.}$$

### 3.6 Instruction Exercises - Sampling Theory

The solutions to these exercises may be found under the page **5ETC0 Canvas Page Modules → Week 1 → B. Sampling Theorem (Ideal sampling case) and Dimensionality Theorem**

**Exercise 1) (Video solution available)** An arbitrary waveform  $w(t)$  with a rectangular spectrum with the highest-frequency  $B = 10$  MHz is sampled with ideal impulses with sampling interval  $T_s$  ("impulse sampling"). The resulting sampled signal is  $w_s(t) = w(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_s)$

- a) What is the maximum value of  $T_s$  with which an exact reconstruction of  $w(t)$  is still possible?
- b) Sketch the spectrum of  $w_s(t)$  for a value of  $T_s = 20$  ns. Clearly show the characteristic points on both axes!
- c) What is the impulse response of the ideal recovery filter?

## 4 Sampling Methods (Pulse Amplitude Modulation)

### 4.1 Learning objectives

Students completing this chapter should have learned:

1. Understand how natural gating and flat top sampling differ from ideal sampling, and why ideal sampling is practically not possible to be implemented
2. Be able to sketch the spectrum and time evolution of a signal sampled with either natural gating or flat top sampling methods.
3. Be able to explain the advantages and disadvantages of the 3 different sampling methods.

### 4.2 Motivation

In the previous chapter, we learned about the sampling theorem and the unique case of ideal sampling. We also saw that ideal sampling is not achievable in practice (due to the fact that the ideal interpolation (reconstruction) function is non-causal, or alternatively that the required frequency domain filter has a rectangular shape with an infinite slopes). This section introduces two methods that can be achieved practically, namely gating (natural sampling) and flat-top sampling. These two methods can be used to sample an analog signal, and they are used in practice, as a step before quantization of the signal which finally gets transformed into the digital domain.

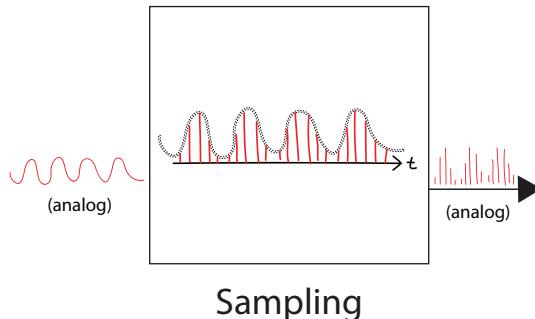


Figure 20: Sampling Methods (Pulse Amplitude Modulation) - Topic map location

### 4.3 Gating (natural sampling)

Natural sampling is a method, which can be best described, where an analog signal  $a(t)$  is allowed to pass through a circuit for a certain amount of time  $\tau$ , followed by a period where the signal is blocked. This sequence repeats itself periodically as can be seen in Fig. 21. Mathematically it is equivalent to multiplying the original signal  $a(t)$  with a train of unit-amplitude rectangular sampling pulses  $x_g(t)$ .

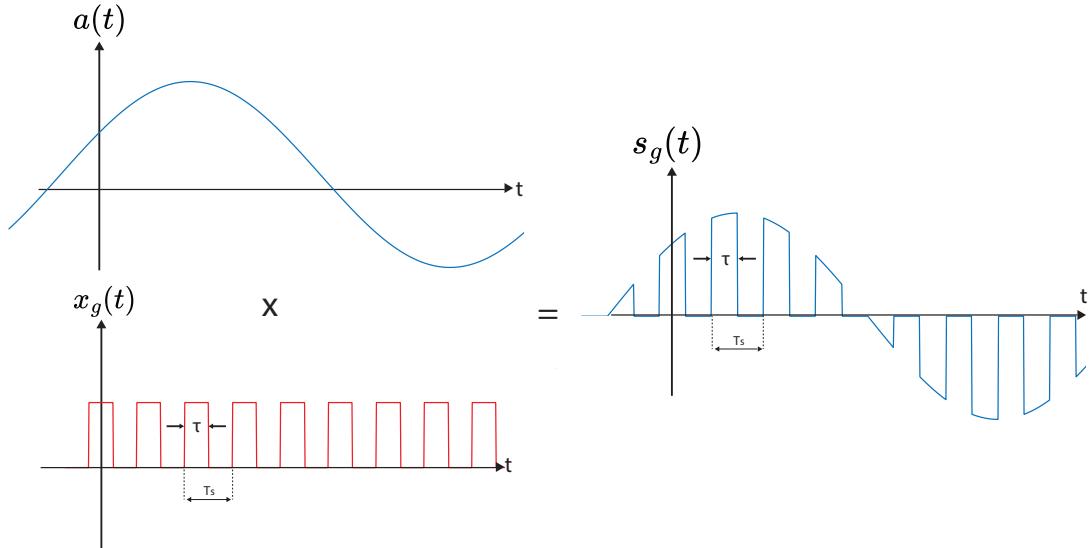


Figure 21: A input signal  $a(t)$  being multiplied by a finite width impulses  $x_g(t)$ , to generate output signal  $s_g(t)$ , illustrating the concept of gating (natural sampling)

This type of sampling can be achieved by a switch and a clock source with duty cycle  $d = \frac{\tau}{T_s}$  and sampling period  $T_s$ . (See Fig. 22)

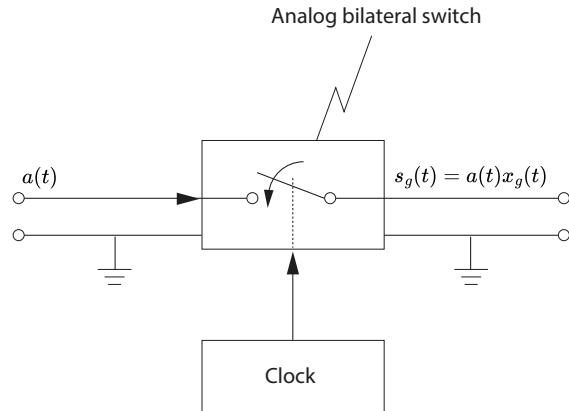


Figure 22: Generation of a sampled waveform using natural sampling (gating). [2, ch.3-2, p.167]

#### 4.3.1 Spectrum of natural sampling

Firstly, let us analyze the spectrum of the finite-width impulses  $x_g(t)$ . Since  $x_g(t)$  is a periodic signal, we can apply the Fourier series.  $x_g(t)$  can be then expressed as

$$x_g(t) = \sum_{k=-\infty}^{\infty} \Pi\left(\frac{t - kT_s}{\tau}\right) \quad (46)$$

where  $\tau$  is the pulse width. Now since  $x_g(t)$  is periodic, the Fourier series expansion is given by

$$x_g(t) = \sum_{n=-\infty}^{\infty} c_n e^{jn\omega_s t} = \sum_{k=-\infty}^{\infty} \Pi\left(\frac{t - kT_s}{\tau}\right) \quad (47)$$

where  $\omega_s = 2\pi f_s$  is the sampling frequency and  $c_n$  are the fourier series coefficients. To find the coefficients we have

$$c_n = \frac{1}{T_s} \int_{T_s} s(t) e^{-j n \omega_s t} dt = \frac{1}{T_s} \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} e^{-j n \omega_s t} dt = \frac{\tau}{T_s} \text{sinc}(n \cdot \frac{\tau}{T_s}) \quad (48)$$

Furthermore,  $X_g(f)$  can be expressed as follows (*using the fact that multiplication by  $e^{j2\pi f_s t}$  in time-domain is a shift in frequency domain  $\delta(f - nf_s)$* )

$$X_g(f) = \mathcal{F}\left(\sum_{n=-\infty}^{\infty} c_n e^{jn\omega_s t}\right) = \sum_{n=-\infty}^{\infty} c_n \delta(f - nf_s) \quad (49)$$

Replacing  $c_n$  on Eq. (49) with that found on Eq. (48) yields the final spectrum of  $X_g(f)$

$$X_g(f) = \sum_{n=-\infty}^{\infty} \frac{\tau}{T_s} \text{sinc}(n \cdot \frac{\tau}{T_s}) \delta(f - nf_s) \quad (50)$$

The output sampled signal  $s_g(t)$  can be expressed as

$$s_g(t) = a(t)x_g(t) = a(t) \cdot \sum_{k=-\infty}^{\infty} \Pi\left(\frac{t - kT_s}{\tau}\right) \quad (51)$$

Finally, the spectrum of the natural sampled output signal  $S_g(f)$  can be derived using the fact that multiplication in the time domain is convolution in the frequency domain as

$$S_g(f) = \mathcal{F}[s_g(t)] = \mathcal{F}[a(t)x_g(t)] = A(f) * X_g(f) \quad (52)$$

Placing Eq. (50) in Eq. (52) we finally get the spectrum of  $S_g(f)$

$$S_g(f) = A(f) * \sum_{n=-\infty}^{\infty} \frac{\tau}{T_s} \text{sinc}(n \cdot \frac{\tau}{T_s}) \delta(f - nf_s) = \sum_{n=-\infty}^{\infty} \frac{\tau}{T_s} \text{sinc}(n \cdot \frac{\tau}{T_s}) A(f - nf_s) \quad (53)$$

$S_g(f) = d \sum_{n=-\infty}^{\infty} \text{sinc}(nd) A(f - nf_s)$

(54)

where  $d$  is the duty-cycle and is expressed as  $d = \frac{\tau}{T_s}$ . In Fig. 23, the analytical spectrum of natural sampling is shown. From the derivations above, and illustration in Fig. 23, It can be observed that the sampled spectrum is a repeated spectrum of  $a(t)$ , as was also shown in the sampling theory section. However, it is also weighed by the Fourier transform of the gating waveform  $x_g(t)$ , which follows a sinc function. The first zero on the bandwidth occurs at  $f = 1/\tau$ .

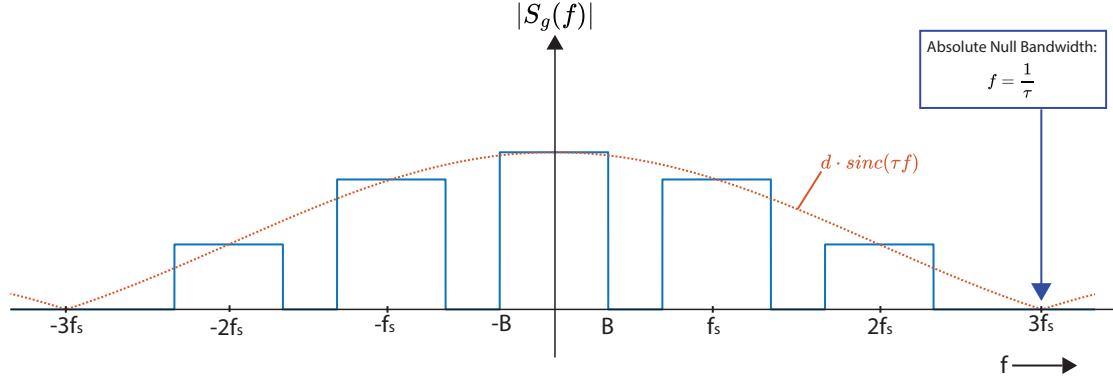


Figure 23: Analytical gating (natural sampling) spectrum for  $d = 1/3$ ,  $f_s = 3$  Hz, assuming that the analog signal  $A(f)$  has a rectangular spectrum from -1 Hz to 1 Hz ( $B=1$  Hz).

#### 4.4 Flat-top sampling (or instantaneous sampling)

In flat-top sampling, the value of the analog signal is captured at the sampling instant  $kT_s$  and held constant for the duration of the sample pulse  $\tau$  (it is hence often referred to as sample and hold method). This concept is illustrated in Fig. 24

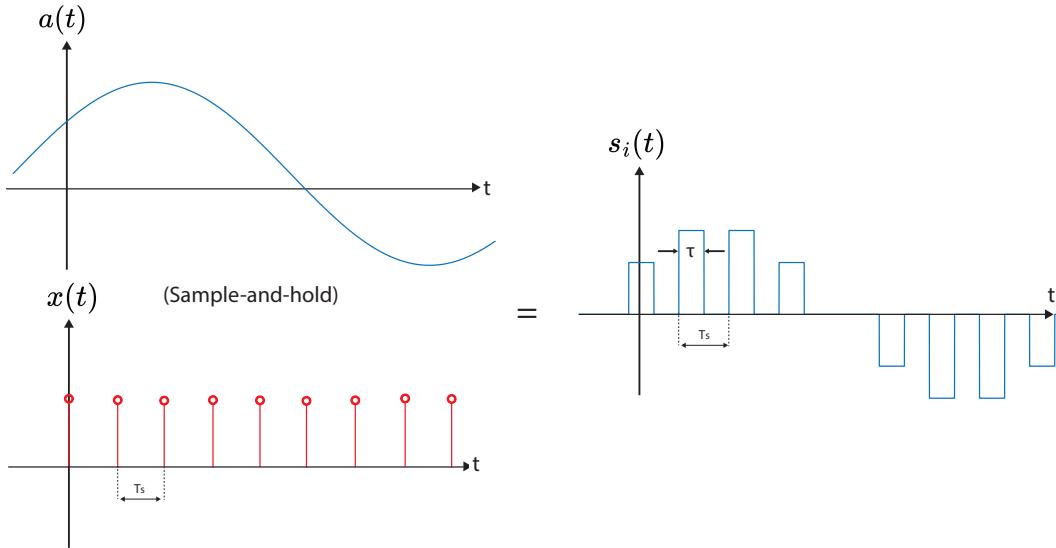


Figure 24: A input signal  $a(t)$  undergoing sample-and-hold to generate output flat-top sampled signal  $s_i(t)$ , illustrating the concept of flat-top (instantaneous) sampling

Unlike gated sampling, which results in a sampled signal with varying amplitude during the sample time, Flat-top sampling produced a single voltage value for the duration of the sampling pulse duration. This makes it a practical method since the fixed value can be easily converted into a digital value (quantized) as will be shown in the following chapters.

##### 4.4.1 Spectrum of flat-top sampling

The spectrum of flat-top sampling can be found more easily by exploiting the structure of  $s_i(t)$ . Mathematically,  $s_i(t)$  using flat-top sampling can be represented as

$$s_i(t) = \sum_{k=-\infty}^{\infty} a(kT_s) \Pi\left(\frac{t - kT_s}{\tau}\right) \quad (55)$$

We can change  $s_i(t)$  to

$$s_i(t) = \sum_{k=-\infty}^{\infty} a(kT_s)\Pi\left(\frac{t}{\tau}\right) * \delta(t - kT_s) \quad (56)$$

$$s_i(t) = \Pi\left(\frac{t}{\tau}\right) * [a(t) \sum_{k=-\infty}^{\infty} \delta(t - kT_s)] \quad (57)$$

By taking the Fourier transform of Eq. (57) we get

$$S_i(f) = \mathcal{F}\{s_i(t)\} = dsinc(\tau f)(\mathcal{F}\{a(t) \sum_{k=-\infty}^{\infty} \delta(t - kT_s)\}) \quad (58)$$

where  $d = \frac{\tau}{T_s}$ , equal to the duty cycle of the sample pulse and the fact that  $\mathcal{F}\{\Pi(\frac{t}{\tau})\} = dsinc(\tau f)$ . We can realize that  $a(t) \sum_{k=-\infty}^{\infty} \delta(t - kT_s)$  is identical to ideal impulse sampling, so we can use the result of Eq. (37), to solve directly.

$$S_i(f) = \mathcal{F}\{s_i(t)\} = dsinc(\tau f)[\frac{1}{T_s} \sum_{k=-\infty}^{\infty} A(f - kf_s)] \quad (59)$$

Thus, finally, we have that the spectrum of flat-top sampling is given as:

$$\boxed{S_i(f) = \frac{1}{T_s} |dsinc(\tau f)| \sum_{k=-\infty}^{\infty} |A(f - kf_s)|} \quad (60)$$

From Eq. (61), we can see that the spectrum of flat-top sampling is a filtered spectrum of ideal impulse sampling by a form of low-pass filter. The spectrum of flat-top sampling is illustrated in Fig. 25. An important observation can be made from Fig. 25, that some high-frequency components might be altered, due to the filtering effect of flat-top sampling. This is known as *aperture effect*. This may be resolved by (i) narrowing the pulse width  $\tau$ , (ii) by an equalization filter with transfer function  $\frac{1}{H(f)}$  where  $H(f)$  is the transfer function of flat-top sampling.

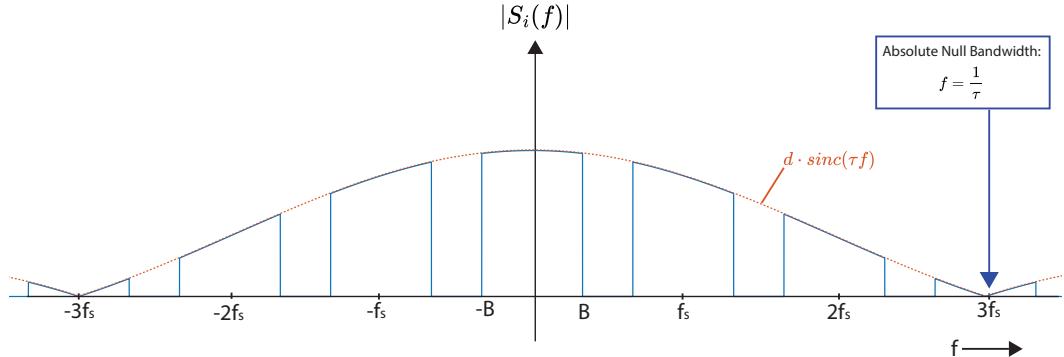
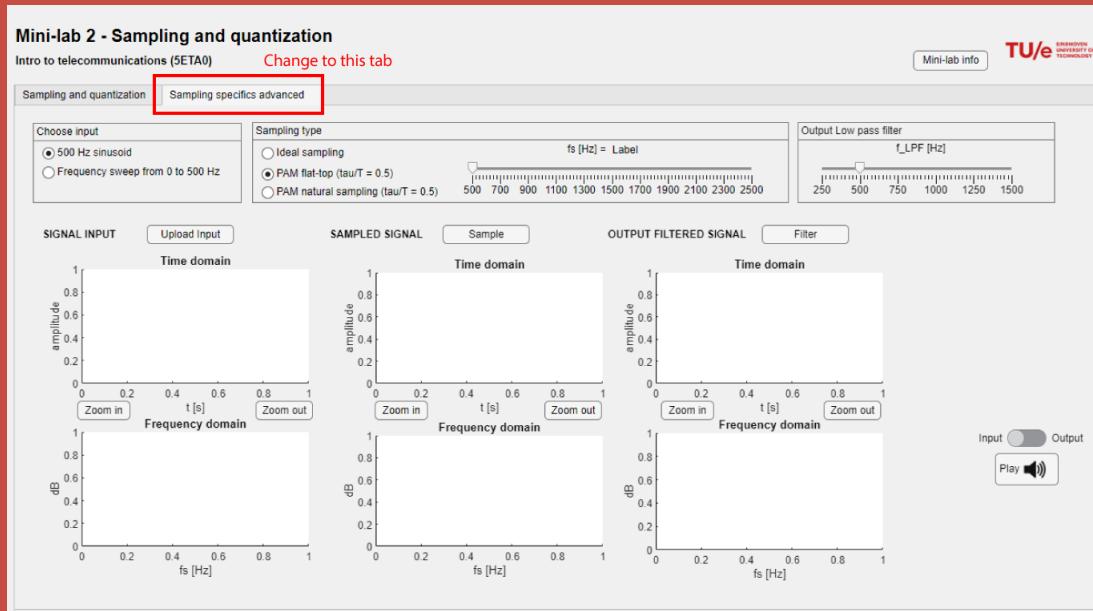


Figure 25: Analytical **flat-top (instantaneous)** sampling spectrum for  $d = 1/3$ ,  $f_s = 3$  Hz, assuming that the analog signal  $A(f)$  has a rectangular spectrum from -1 Hz to 1 Hz ( $B=1$  Hz).

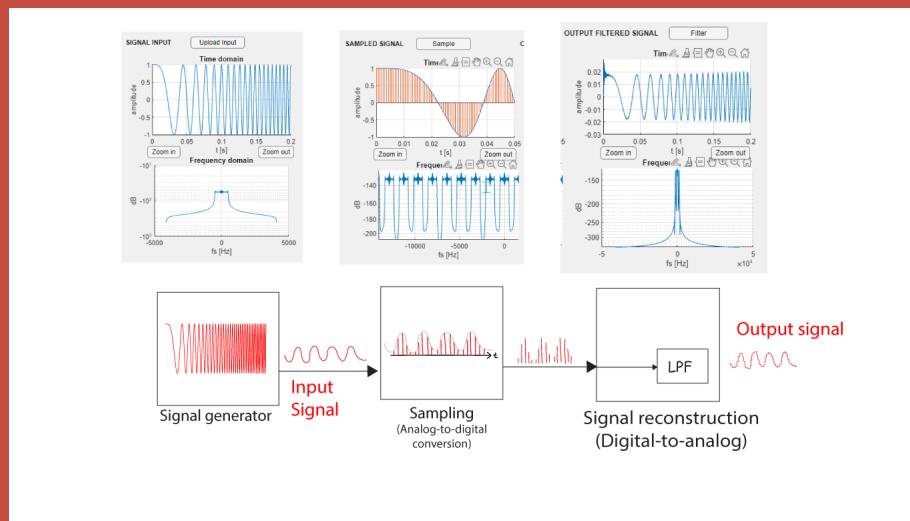
## Minilab exercise 4.1 - Gating and flat-top sampling demonstration

This mini-lab exercise requires you to use Mini-lab 2 (Sampling and Quantization) on MATLAB. (See section 1.3 of the reader in case you dont know how to install and start)

When you open Minilab 2, please click on the 'sampling specifics' and you will be confronted with the following view:



The figure below explains the signal chain being considered in this Minilab exercise. The main purpose is to test different sampling methods, and try to reconstruct them at the output using a low-pass filter. We don't consider any other part of the digital communication chain, and the DAC (Digital to analog converter) is removed from the signal reconstruction block, since we are analyzing only the output of sampling, which is on the analog domain.



Change the input signal to a frequency sweep from 0 to 500 hz. This type of signal, known as a *chirp signal*, it has a rectangular spectrum and is widely used in radar systems. Press on the button *upload input* to upload the signal, and select *input* and play audio, if you wish to hear how such a *chirp* signal sounds. You may choose the sampling type and the sampling frequency  $f_s$  on the sampling type section. And finally, you may change the passband of the lowpass filter using the knob in the output low-pass filter section.

- 1) For the first demonstration, make sure the input is 500 Hz sinusoid, use ideal sampling, keep the sampling frequency at 500 Hz, and the output low-pass filter knob at 1000 Hz. Upload the input, sample it, and filter it. Play the input of the 500 Hz, and then play the output of the 500 Hz. Do they sound different? If so, what is happening with them? (What is this effect called?). Can you show on the frequency domain of the sampled signal the issue (hint: see sampling theory section) ?
- 2) Now let's move on to more practical sampling methodologies, switch the input type to a chirp signal (*frequency sweep from 0 to 500 Hz*), and make the sampling type *PAM flat-top*. Your task here is to choose an appropriate sampling frequency and tune the output lowpass filter, such that the output *chirp* sounds similar to the input *chirp*. If you get the correct output sound, then move to the next task
- 3) Now that you have correctly tuned everything, make sure that the sampling frequency is smaller than 1200 Hz. Set the lowpass filter to 1500 Hz, sample again and filter, and play the sound of input and output again. Do they sound similar? If no, can you observe what is happening on the spectrum of the output-filtered signal? Try to intuitively explain these effects.
- 4) Now choose *PAM natural sampling* as your sampling method and redo steps 2 and 3, do you notice any difference with the previous?

## 4.5 Instruction Exercises - Sampling methods

The solutions to these exercises may be found under the page **5ETC0 Canvas Page Modules** → **Week 1** → **C. Sampling Methods (Pulse Amplitude Modulation)**

**Exercise 2 (Video solution available)** An arbitrary waveform  $w(t)$  with a rectangular spectrum with the highest-frequency  $B = 10 \text{ MHz}$  is sampled with a rectangular signal with sampling interval  $T_s = 20 \text{ ns}$  ("natural sampling" or "gating"). The rectangular sampling signal has a pulse duration  $\tau = T_s/3$

- a) Make a sketch of the PAM output signal as a function of time.
- b) Draw the spectrum of the PAM output signal. Clearly indicate the characteristic points and add an explanatory text.

**Exercise 3 (Video solution available)** With a "sample and hold" A/D converter an analog input signal is converted into a "flattopped" PAM signal. The sampling frequency is  $25 \text{ kHz}$ ; the width of the PAM pulses is  $10 \mu\text{s}$

- a) What is the maximum frequency that can be reliably transmitted with this system?

A music signal  $w(t)$  with maximum frequency  $f_x = 7 \text{ kHz}$  is applied to this system.

- Draw the PAM output signal as a function of time, and draw the corresponding spectrum — $W(f)$ —. Clearly indicate the characteristic points on the axes!

## 5 Digitization

### 5.1 Learning objectives

Students completing this chapter should have learned:

1. Can derive the bit rate, symbol rate and bandwidth requirements for a digital transmission system based on the original analog signal and the sampling of it.
2. Understand the impact of the number of quantization levels and the limiting factors for quantization.
3. Understand and be able to indicate which noise mechanism (quantization or bit error rate) dominates the SNR of the reconstructed signal in a digital communication system.
4. Can calculate the required  $B_{pcm}$  bandwidth for a digitized signal with  $n$  quantization bits.
5. Can find/calculate the signal to noise ratio ( $SNR_{in}$ ) at the entrance of a receiver based on the required Bit Error Rate (BER)
6. Can calculate the BER based on the  $SNR_{in}$  using the Q function.
7. Understand the fact that increasing or decreasing the number of quantization bits does not always lead to a respective increase or decrease in  $SNR_{out}$  performance.

### 5.2 Motivation

In this chapter, we transition from understanding the sampling of analog signals to the conversion of these samples into a digital binary stream. Pulse Code Modulation (PCM) lies at the heart of this process, transforming analog signals into discrete digital streams, which is at the core of digital communications. As you study PCM, you will explore crucial concepts like bit rate, bits per sample, spectral efficiency and signal-to-noise ratio illuminating their impact on bandwidth utilization.

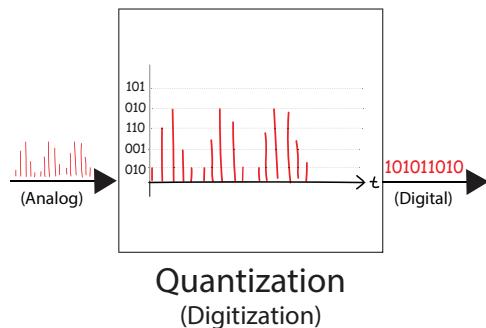


Figure 26: Digitization - Topic map location

### 5.3 Pulse code modulation (PCM)

Pulse code modulation (PCM) is the name for the process by which the sampled analog values of the signal, one wants to digitize, are converted into a stream of bits.

#### 5.3.1 From samples to quantized information

The PCM signal is generated by converting the samples obtained from the sampling circuit into a stream of bits. In the previous sections, we have discussed in detail the process of

sampling and sampling methods. In this section, we discuss quantization in detail and its performance. We will assume throughout this section that we have an incoming flat-top sampled signal from the sampling stage, which is still analog, and now we need to convert it to a digital bit stream. This conversion is done through *quantization*. From the name, this means actually transforming the samples into a sequence of bits (thus we make an analog-to-digital conversion (ADC)). In hardware, this implies an ADC circuit will be used (the electronic and circuit aspects of ADC are beyond the scope of this course and are handled in other courses in the EE curriculum).

During quantization, we start by separating a voltage range (for ex -5 V to 7 V) into  $M$  levels, and we assign a binary codeword to each of the levels. Then, we can represent any voltage level with a binary code word by rounding it to the nearest level, and in this way, we essentially convert the analog samples into digital codewords of  $n$  bits. This concept is illustrated in Fig. 27. We may note that the **number of quantization levels  $M$**  is expressed as

$$M = 2^n \quad (61)$$

where  $n$  is the number of bits per sample. Furthermore, the parallel bits obtained by the ADC circuit are converted into a digital binary stream as can be observed from the figure. The output of the quantizer circuit (ADC) is then further processed (encoded, shaped and often processed to support error correction and detection) to finally be transmitted through the channel. All of these aspects will be discussed in the following sections.

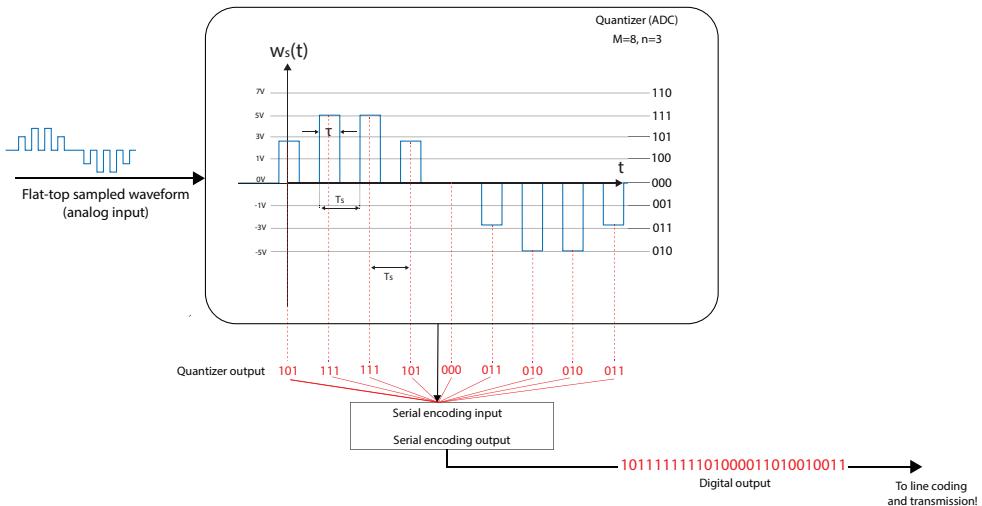


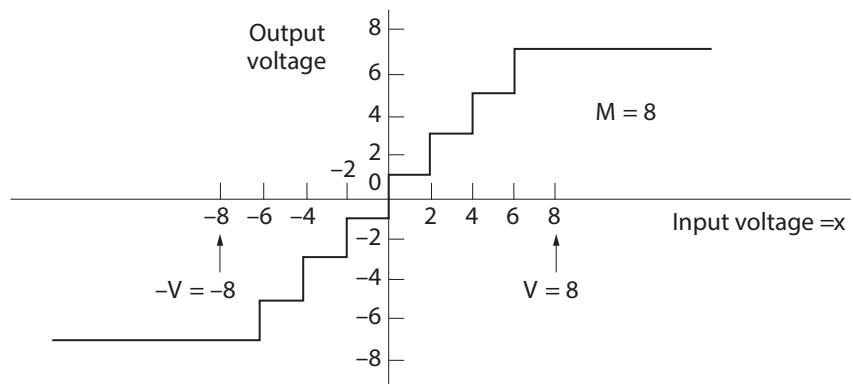
Figure 27: Illustration of quantization and encoding, with  $M = 8$  and  $n = 3$ .

The following subsections, explain the concepts of bit-rate  $R$  and spectral efficiency, which are a pre-knowledge to understand the PCM bandwidth requirements.

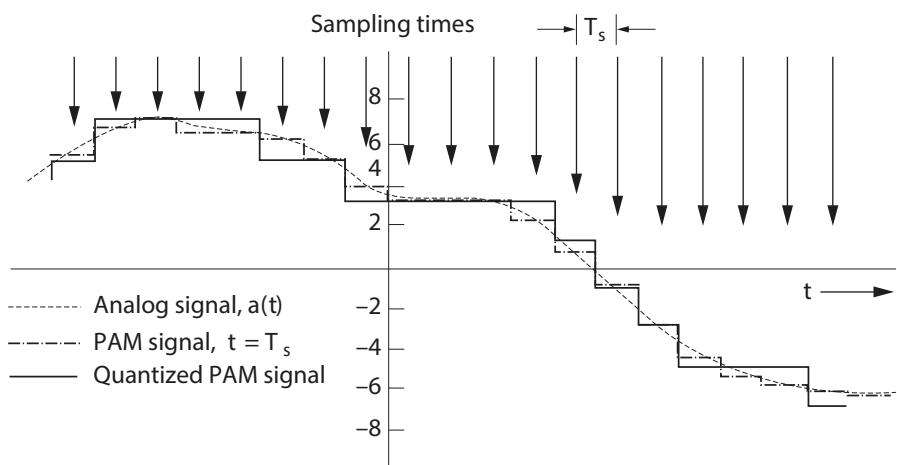
### 5.3.2 Bit-rate $R$

Assuming we sample an analog signal a certain amount of times per second  $f_s$ , and quantize and finally transmit that information (the level of the signal at each point) with  $n$  bits, we deduce that the bit rate ( $R$ ) will be the multiplication of the two, hence:

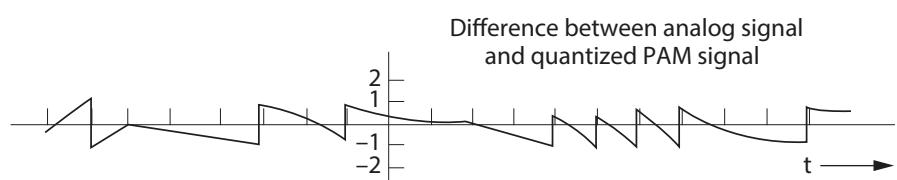
$$R = n \cdot f_s \quad [\text{bits/s}] \quad (62)$$



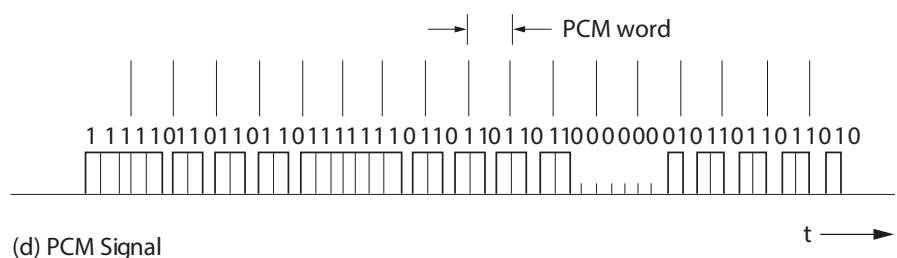
### (a) Quantizer Output-Input Characteristics



(b) Analog Signal, Flat-top PAM Signal, and Quantized PAM Signal



(c) Error Signal



(d) PCM Signal

Figure 28: Illustration of waveforms in a PCM system [3]

### 5.3.3 Spectral efficiency $\eta$

The spectral efficiency of a digital signal is given by the number of bits per second of data that can be supported by each hertz of bandwidth. That is,

$$\eta = \frac{R}{B} \text{ (bits/s)/Hz} \quad (63)$$

Where R is the bit rate (or data rate) and B is the bandwidth of a PCM signal. The bit rate R, can be directly derived from the sampling frequency and number of quantization bits (see Eq. 62). The bandwidth B, needed for successfully sending the information with a bit rate R, is determined by the type of line coding used and the number of bits included in each symbol (this will be elaborated upon later in this reader). Hence it is sometimes also valuable to consider efficiency in terms of symbols/sec/Hz. If symbol rate (D) is known, and the spectral properties of the chosen line coding are known (shape and number of bits per symbol), one can deduce the required bandwidth for successful transmission. We derive in chapter 6 a lower bound for the required bandwidth for a communication channel based on the symbol rate D.

### 5.3.4 PCM Bandwidth

Since we want to transmit PCM signals (binary waveforms), the bandwidth of binary PCM waveforms is important. Because most channels have bandwidth limitations, we need to make sure that the bandwidth of the PCM signal is supported inside the communication channel, hence indicating the importance of analyzing the PCM bandwidth  $B_{pcm}$ . If the channel bandwidth is smaller than the bandwidth of a PCM signal, the signal will be filtered. If that happens then the binary pulses will be smeared into neighboring bit slots, causing *intersymbol-interference (ISI)*, which is highly unwanted as it will corrupt the data as will be discussed in detail later.

The PCM bandwidth depends on the bit rate and the waveform pulse shape used to represent the binary data (line coding). For no aliasing, we require that  $f_s \geq 2B$  where B is the bandwidth of the original analog signal being digitized. From the dimensionality theorem, it can be shown that the bandwidth of a binary encoded PCM waveform is bounded by

$$B_{pcm} \geq \frac{1}{2}R = \frac{1}{2}nf_s \geq nB \quad (64)$$

This minimum PCM bandwidth of  $\frac{1}{2}R = \frac{1}{2}nf_s$  is obtained only when sinc pulse shapes are used to generate the PCM waveform, instead of rectangular waveforms. Nevertheless, usually, a more rectangular type of pulse shape is used (which will be discussed later in the section dealing with line codes), and consequently, the PCM bandwidth will be larger than this minimum. We can also see from Eq. (64), that PCM requires a bandwidth at least n times as large than that of the input analog signal. A general equation for the bandwidth usage of PCM signals can be formulated as

$$B_{pcm} = \frac{1}{\eta}nf_s \quad (65)$$

where  $\eta$  is the spectral efficiency of the line coding, n the number of bits per sample and  $f_s$  the sampling frequency.

## 5.4 Signal-to-noise ratio (*after signal reconstruction*)

Once a signal has been digitized we would like to estimate the resulting quality of the reconstructed signal after reception. A typical way to estimate the quality of a signal is to evaluate the ratio between signal power and noise. This is often referred to as the Signal-to-noise ratio (SNR). Specifically in the case of signal reconstruction, we will define the  $SNR_{out}$  as the relationship between the original analog signal before digitization and the reconstructed waveform at the output of the receiver. Any error in the reconstruction, and hence a deviation from original signal, will be considered as noise. We distinguish between two main mechanisms that may result in the reconstructed signal deviating from the original signal:

- **Quantising errors** - Caused by round-off errors in the approximating M-step quantizer with respect to the original analog sampled signal. *This is actually happening already at the transmitter side and is inherent to the process of conversion from an analog to a digital signal!*
- **Bit errors in the recovered PCM signal** - Caused by channel noise and intersymbol interference due to improper channel filtering. *This error mechanism is caused often by the channel or the receiver circuitry.*

We can easily illustrate the effect additive noise can have on the detection of the transmitted digital signal in Figure 29. Here we can see that depending on the amount of additive noise in the channel, and the resulting SNR at the input of the receiver, the chance of bits being wrongly detected may increase. For the case of SNR of 60 dB the signal power is much bigger than noise power, and the signal can be seen clearly. For 20 dB the noise can be seen, however the signal can be easily differentiated from the noise. On the other hand, for 0 dB the signal power and noise power are equal, so the signal information is lost in the noise. As will be discussed later in this chapter, the lower the SNR is, the higher the likelihood of a bit being detected wrongly.

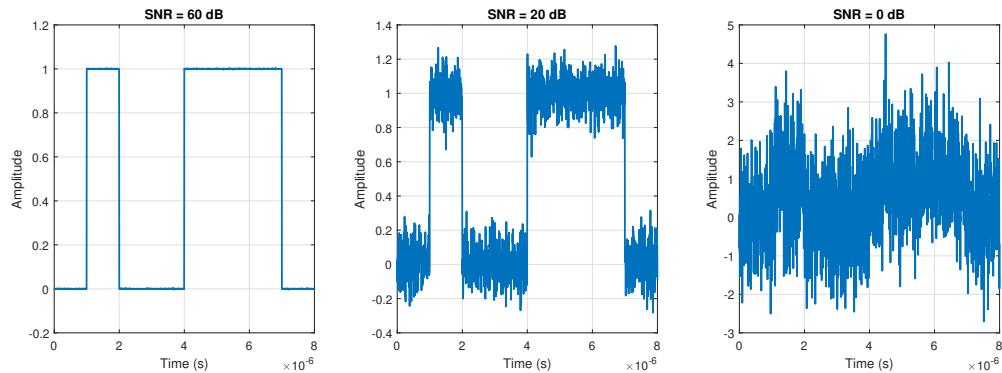


Figure 29: Illustration of different levels of signal-to-noise ratio (SNR) on a digital binary signal

### 5.4.1 Receiver output signal-to-noise ratio

Under certain assumptions, it can be derived that the signal-to-noise ratio at the output of the receiver, is expressed as

$$\left(\frac{S}{N}\right)_{out} = \frac{M^2}{1+4P_e(M^2-1)} \quad (66)$$

where  $P_e$  is the bit error probability (discussed in detail later) and  $M$  is the number of quantization levels. As discussed above, SNR at the output of the receiver depends on the level of quantization and bit error effects. Furthermore, the peak signal-to-noise ratio at the output of the receiver is expressed as

$$\left(\frac{S}{N}\right)_{pk,out} = \frac{3M^2}{1+4P_e(M^2-1)} \quad (67)$$

If,  $P_e \approx 0$ , then we may assume that the SNR is only based on quantization noise which then leads to

$$\left(\frac{S}{N}\right)_{out} = M^2 \quad (68)$$

and for peak SNR being

$$\left(\frac{S}{N}\right)_{pk,out} = 3M^2 \quad (69)$$

Recalling that  $M = 2^n$ , we may express Eqs. (68)-(69) in decibels as

$$\left(\frac{S}{N}\right)_{out,dB} = 6.02n + \alpha \text{ [dB]} \quad (70)$$

where  $n$  is the number of bits per sample and  $\alpha = 4.77$  dB for the peak SNR and  $\alpha = 0$  for the average SNR. This equation is called the **6-dB rule**, which points out a significant characteristic for PCM: *An additional 6-dB improvement in  $SNR_{out}$  is obtained for each bit per sample added to the PCM word. This holds true as long as  $P_e \approx 0$  and only if the distribution of signal is uniform over interval  $-V$  to  $+V$ .* The performance of the PCM system for the most optimistic case (i.e.,  $P_e = 0$ ) can easily be obtained as a function of  $M$ , to show the 6dB improvement rule. This result is shown in Fig. 46.

TABLE 3-2 PERFORMANCE OF A PCM SYSTEM WITH UNIFORM QUANTIZING AND NO CHANNEL NOISE

Number of Quantizer Levels Used, M	Length of the PCM Word, n (bits)	Bandwidth of PCM Signal (First Null Bandwidth) <sup>a</sup>	Recovered Analog Signal-Power-to-Quantizing-Noise Power Ratios (dB)	
			(S/N) <sub>pk out</sub>	(S/N) <sub>out</sub>
2	1	2B	10.8	6.0
4	2	4B	16.8	12.0
8	3	6B	22.8	18.1
16	4	8B	28.9	24.1
32	5	10B	34.9	30.1
64	6	12B	40.9	36.1
128	7	14B	46.9	42.1
256	8	16B	52.9	48.2
512	9	18B	59.0	54.2
1,024	10	20B	65.0	60.2
2,048	11	22B	71.0	66.2
4,096	12	24B	77.0	72.2
8,192	13	26B	83.0	78.3
16,384	14	28B	89.1	84.3
32,768	15	30B	95.1	90.3
65,536	16	32B	101.1	96.3

<sup>a</sup> B is the absolute bandwidth of the input analog signal.

Figure 30: Table of S/N ratios

### Overview of quantization noise types (*extra material*).

The quantizing noise at the output of the PCM receiver can be categorized into four types:

- 1) **Overload noise** - If an input analog signal, exceeds the ADC range, then flat tops start occurring in the recovered signal because the ADC saturates.
- 2) **Random noise** quantization causing round-off errors, for normal input signals within the ADC range (see Fig. 28c).
- 3) **Granular noise** - for signals relatively small in comparison to the quantization levels (decreases with more quantizing levels, or nonlinear quantization). See Fig. 31, for illustration
- 4) **Hunting noise** - for nearly-constant input signals, may cause oscillating quantizer.

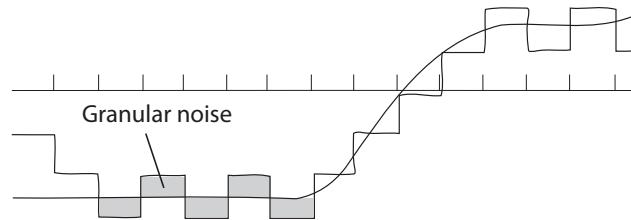
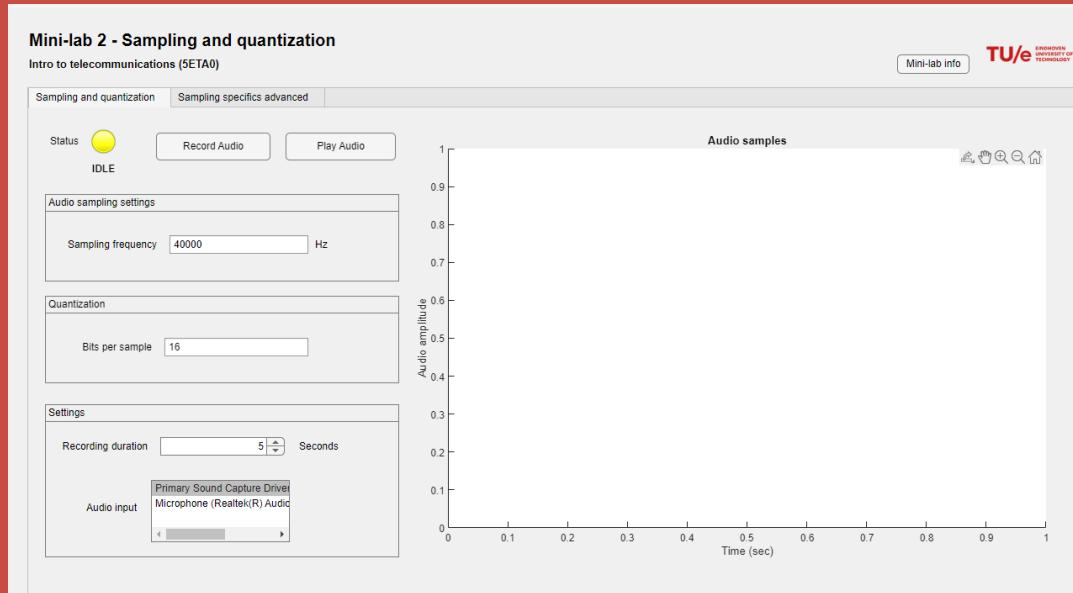


Figure 31: Granular noise in quantization [2, ch.3, p.200]

## Minilab exercise 5.1 - Quantization levels and noise

This mini-lab exercise requires you to use Mini-lab 2 (Sampling and Quantization) on MATLAB.

When you open Minilab 2 you will be confronted with the following view:



In the settings section, you may change the microphone recording duration and also which microphone to use from your computer. Furthermore, on the *audio sampling settings* you may change the sampling frequency, and for this exercise, that is of interest. In this minilab, we are interested in the settings of the number of bits per sample. We may note, that you can only choose bits per sample to be 8, 16 or 24

- 1) The task in this minilab is to simply experience the effect of having a low number of bits per sample, and a high number of bits per sample, and what happens to the audio. Start by setting the number of bits per sample to 24, record your audio and play it.
- 2) Now, try to make the same sound when you recorded the first step, but change the number of bits per sample to 16 and record your voice again. Do you see or hear any difference?
- 3) Record the same sounds but with 8 bits per sample. Notice any differences in the audio? And in the audio samples plot?

## 5.5 Bit-error probability $P_e$ and channel noise

Channel noise refers to unwanted signals or interference that disrupt the transmission of information through a communication channel. This interference can arise from various sources, including thermal noise, electromagnetic interference, and crosstalk from adjacent channels. Any physical channel might add some form of noise to a signal passing through. The goal is to correctly detect the transmitted bits on the incoming noisy signal while making a few mistakes in detection as possible.

Bit-error probability ( $P_e$ ) is a crucial metric in communication systems, representing the likelihood of an error occurring in the transmission of a single bit across a communication channel. As an example, an accepted bit-error probability rate for digital wireless communications is around  $P_e \approx 10^{-6}$ . This means that for every 1 million bits, at least one bit will be flipped (hence corrupting the data), which might seem at first as not a big deal, however, if you are downloading data at 1 Mbps speed (which nowadays is considered not that fast), this means that every second one bit will be flipped, and corrupt the data.

To understand how bit-error probability is modeled on a channel, it is firstly important to understand some basic concepts in the field of probability, such as probability density functions, gaussian distributions, and Q-functions.

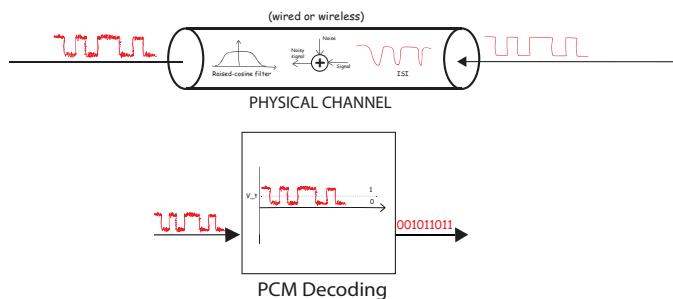


Figure 32: Bit-error probability  $P_e$  and channel noise - Topic map location

### 5.5.1 Probability and random variables

Probability theory is widely used in various fields such as statistics, science, finance, and engineering to analyze uncertainty, make predictions, and make informed decisions based on available information. In digital communications, probability theory is used to model channel uncertainties because if you transmit a signal, you don't know how the signal will be received by the receiver, and the receiver on the other hand, does not know what kind of bits are going to arrive, but the receiver needs to correctly decode your message, so there is a lot of uncertainties to what happens when your signal gets transmitted.

Probability is a mathematical concept that quantifies the likelihood of an event occurring. It is expressed as a number between 0 and 1, where 0 indicates impossibility (the event will not occur), and 1 indicates certainty (the event will occur).

A **random variable** is a mathematical formalization of a quantity or object that depends on random events. For example, if you are designing a digital receiver, and you want to model the chance of receiving a 0 or 1 from the transmitter, then your random variable **X** is a **discrete random variable** that either takes the value 0 or 1.

Furthermore, suppose instead of modeling whether you receive a 0 or 1, you want to model the voltage of the signal you are receiving. Voltage is a continuous quantity, which means that your **random variable X will be continuous** and can take any value in the range  $V_{min} < x < V_{max}$

When we conduct experiments or studies involving random variables, we observe particular outcomes. For example, if we roll a fair six-sided die, the possible outcomes (observations) are the numbers 1 through 6. Each roll of the die results in a specific observation.

### 5.5.2 Probability density functions (PDFs)

The PDF represents the probability distribution of a continuous random variable over its entire range of possible values. It does not directly give the probability of specific outcomes, as the probability of any single point in a continuous distribution is typically zero. Instead, the PDF specifies the relative likelihood of different outcomes occurring.

In our case where we wanted to model the received voltage random variable on the range  $V_{min} < x < V_{max}$ , if each voltage point on that range has an equal chance (or probability of occurring) then the PDF  $p_X(x)$  of that random variable  $\mathbf{X}$  would be in the form of Fig.33

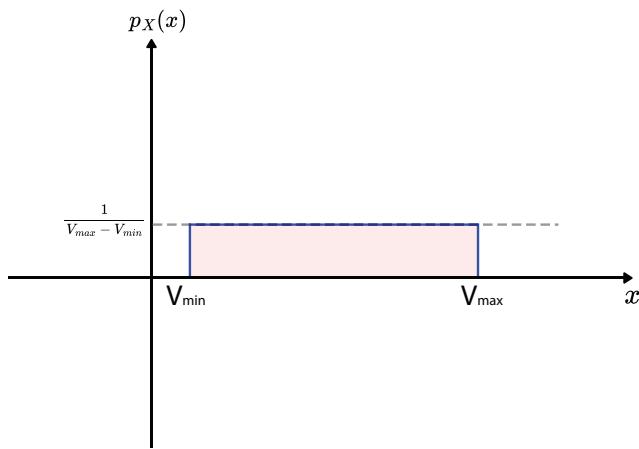


Figure 33: Probability density function of a received voltage random variable with equal probability between range  $V_{min}$  up to  $V_{max}$ . This is also known as a uniform distribution

Now, suppose you want to find the probability that the voltage will fall in a range  $a, b$  within  $V_{min}, V_{max}$ , then

$$\Pr[a < x < b] = \int_a^b p_X(x) dx \quad (71)$$

Evaluating Eq. (71), yields

$$\Pr[a < x < b] = \int_a^b p_X(x) dx = \int_a^b \frac{1}{V_{max} - V_{min}} dx = \frac{(b - a)}{V_{max} - V_{min}} \quad (72)$$

Visually, it means we are summing the area under the PDF between points  $a$  and  $b$  as can be seen in Fig. 34. Most random variables have characteristic PDFs which are called 'distributions' and come from a distribution family. Examples of these families are Binomial distributions, Gaussian distribution, exponential distributions (used to model Queues)

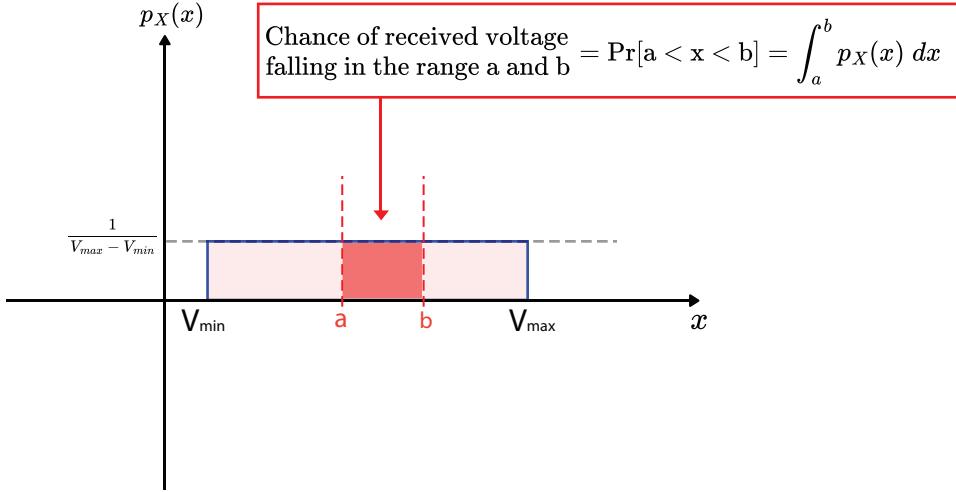


Figure 34: Illustration of finding the probability that the voltage will fall within the range a and b

The following are some properties of PDFs:

For any PDF it must hold that

$$\int_{-\infty}^{\infty} p_X(x) dx = 1 \quad (73)$$

**The average (or mean) value** of a PDF can be found as

$$\mu_X = \int_{-\infty}^{\infty} x p_X(x) dx \quad (74)$$

**The variance** of a PDF can be found as

$$\sigma_X^2 = \text{mean}((x - \mu_x)^2) = \int_{-\infty}^{\infty} (x - \mu_x)^2 p_X(x) dx = \int_{-\infty}^{\infty} x^2 p_X(x) dx - \mu_x^2 \quad (75)$$

**The standard deviation** of a PDF can be found as

$$\sigma_X = \sqrt{\text{Variance}} = \sqrt{\sigma_X^2} \quad (76)$$

Standard deviation measures the 'spreadness' of a PDF.

### 5.5.3 Gaussian distribution (or normal distribution)

Normal distributions are important in statistics and they often emerge naturally in nature. Their importance is partly also due to the central limit theorem. It states that, under some conditions, the average of many samples (observations) of a random variable with finite mean and variance is itself a random variable—whose distribution converges to a normal distribution as the number of samples increases. Therefore, physical quantities that are expected to be the sum of many independent processes, such as measurement errors, often have nearly normal distributions [3].

The PDF of a normal distribution is given by the equation

$$p_X(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \quad (77)$$

where  $\mu_x$  is the mean of the random variable and  $\sigma_x^2$  is the variance of the distribution. Figure 35, illustrates what a Gaussian distribution looks like.

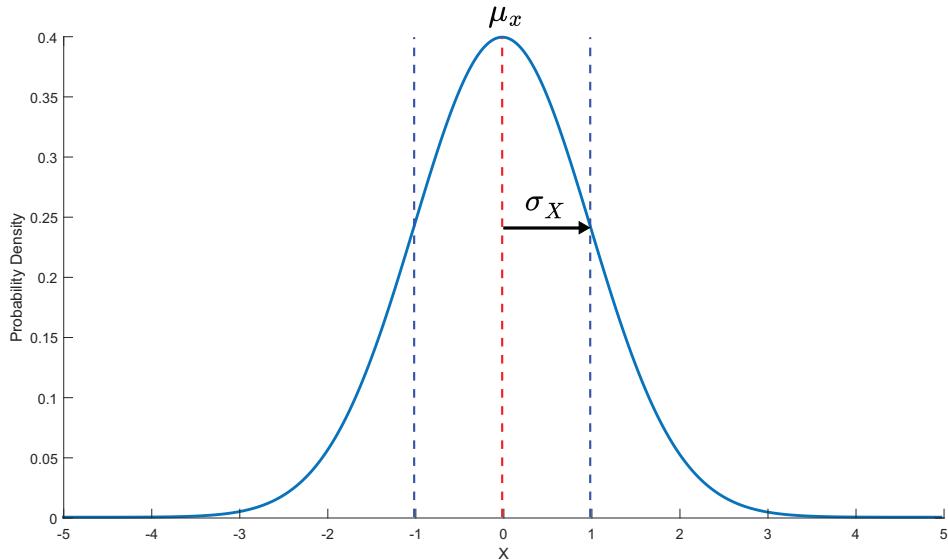


Figure 35: Gaussian distribution with mean  $\mu_x = 0$ , and standard deviation  $\sigma_x = 1$

#### 5.5.4 Additive white Gaussian noise (AWGN)

Additive white Gaussian noise (AWGN) is a basic noise model used in telecommunications to mimic the effects of many random processes that occur in nature that may interfere with your transmitted signal. For example, thermal noise which is generated by the agitation of the charge carriers inside an electrical conductor at equilibrium, is modeled using AWGN noise.

- **Additive** - Because the noise can be added to any signal such that the signal and the noise are combined together, and what we observe or measure is the sum of the two.
- **White** - "white" refers to the spectral characteristics of the noise. White noise has a constant power spectral density across all frequencies denoted by  $\frac{N_0}{2}$ , meaning it has equal power at all frequencies (see Fig. 36).

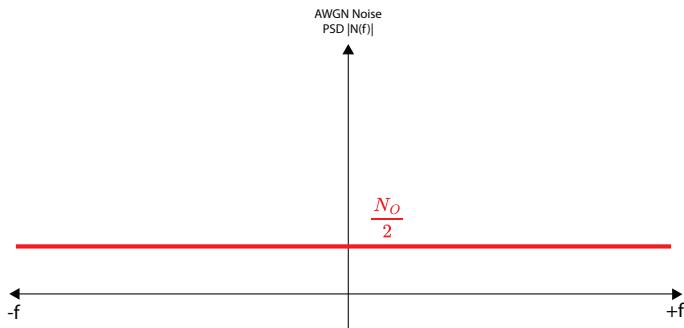


Figure 36: Illustration of AWGN noise power spectral density

- **Gaussian** - This refers to the probability distribution of the noise amplitude. Gaussian noise follows a Gaussian (or normal) distribution, with the mean  $\mu_x = 0$ . Where the square of the standard deviation (called Variance) gives the power spectral density of the noise. Fig. 37 visualizes the relation of AWGN noise in time-domain with the Gaussian distribution.

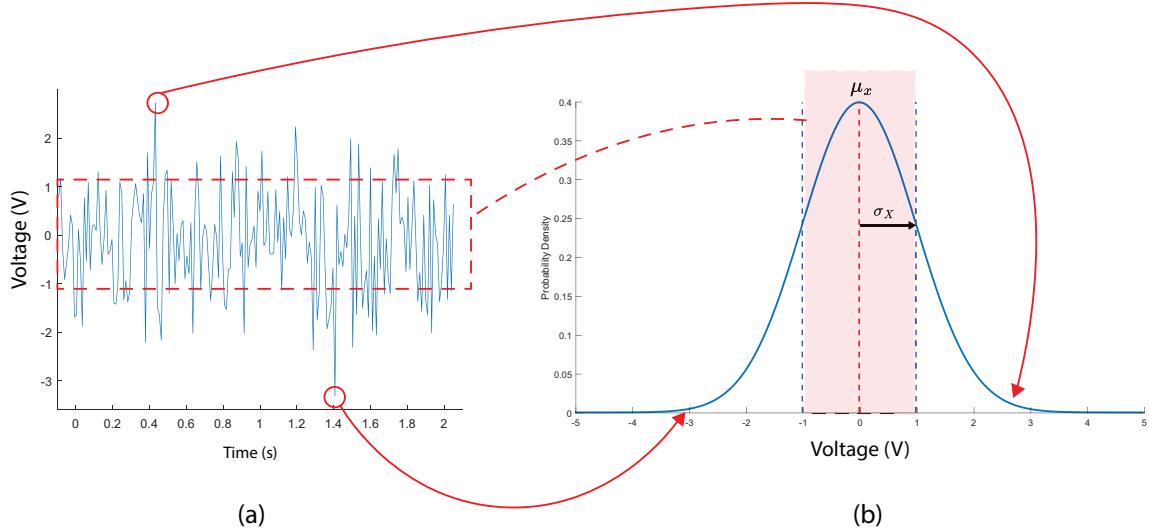


Figure 37: Figure illustrating the relation between Gaussian probability distribution with mean 0 and standard deviation of 1, to AWGN noise in time domain.

We define  $N_O$  as the noise power spectral density, which is a value showing the density of power per hertz. Because of the negative frequencies, we divide  $N_O$  by 2. The total AWGN noise power  $N$ , which is also the variance of the noise distribution, can be computed as

$$N = \sigma^2 = \left(\frac{N_O}{2}\right)(2B) = N_O B \quad (78)$$

To understand why this is the case, it can be visualized in Fig. 38

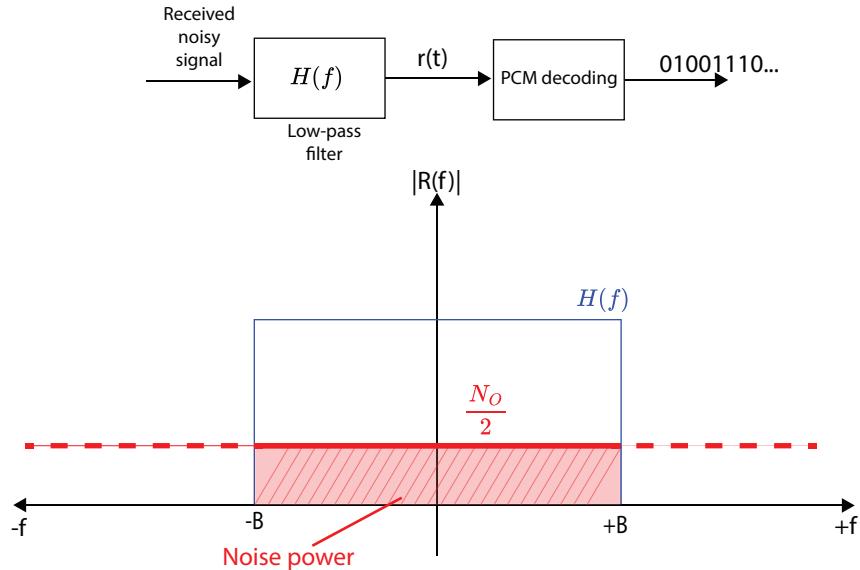


Figure 38: Noise power at the receiver visualization

Every receiver will have a low-pass filter to remove high-frequency noises that are beyond the bandwidth of the signal, hence the AWGN noise power is found by the noise spectral density multiplied by the bandwidth of the transmitted signal (*Which is multiplied by 2, since we also have to include negative frequencies*).

### 5.5.5 Q-function

The Q-function, denoted as  $Q(z)$ , is defined as the tail probability of the Gaussian distribution. Intuitively, it means what is the probability that a random variable with a Gaussian distribution, will take a value larger than  $z$  standard deviations. Q-function mathematically is defined as

$$Q(z) = \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-\frac{u^2}{2}} du \quad (79)$$

where  $z$  is defined as

$$z = \frac{x_0 - \mu}{\sigma} \quad (80)$$

Visually, it means the area of the tail of the Gaussian distribution, which is visualized in Fig. 39

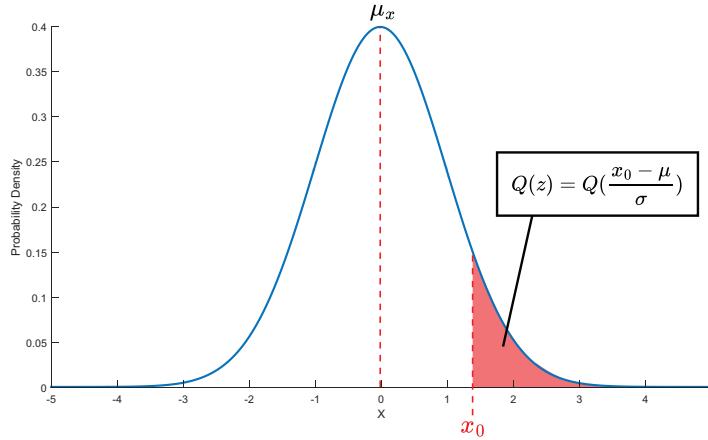


Figure 39: Illustration of Q-function

Nevertheless, you may always use Fig. 40 to get the value of the Q-function based on  $z$  or the other way around (getting the value of  $z$  based on the Q-function result).

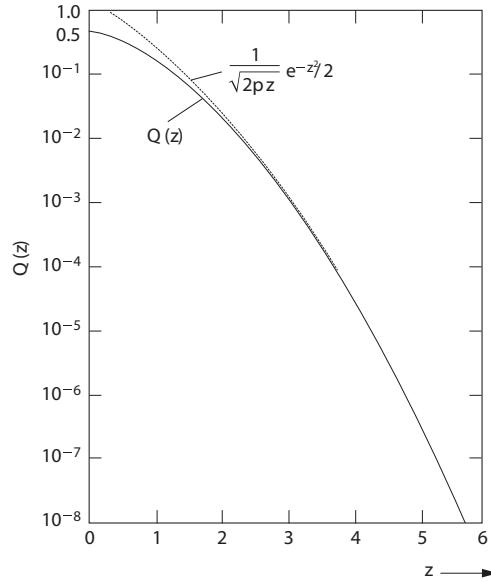


Figure 40: Plot of Q-function upper bounded by  $\frac{1}{\sqrt{2\pi}z} e^{-z^2/2}$ . [2, ch. B-7, p.703]

### 5.5.6 Computing Bit-error probability on AWGN channels

Now that we have covered simple probability theory, q-function, and AWGN, we may finally formulate a method to compute the bit-error probability  $P_e$  of AWGN channels. Suppose we have a channel that is affected by AWGN noise and you want to transmit the binary sequence '01001110. At the receiver, the signal will arrive noisy, and to decode the message, you may define a threshold voltage at 0.5V between 1V and 0V, and any voltage higher than the threshold voltage may be considered as the message '1' and any voltage lower than the threshold voltage may be considered as message '0'. See Fig. 41 for visualizations.

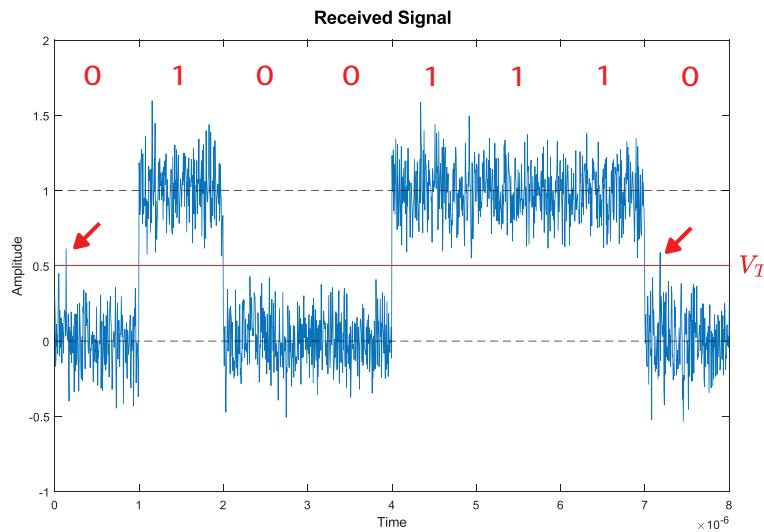


Figure 41: Received noisy signal with binary message 01001110, and a threshold voltage  $V_T$  in the center to decode the message

However, if you look closely at the image you may spot two cases (see the red arrows inside the plot Fig. 41), where the message was 0, however because of noise, the voltage of the signal spiked higher than the threshold voltage  $V_T$ . This is considered an error due to noise because it can cause the decoder to wrongly decode the message (0 instead of 1, or vice versa). Now to quantify this, we want to know what is the chance (or probability) that these spikes will happen, that may cause the receiver to wrongly decode the message.

If we view this problem from the perspective of the probability, given the noise variance  $\sigma^2$ , and the fact that the noise is AWGN noise, then each message becomes a Gaussian distribution with a mean around the original voltage of the message (for example, if voltage 5V represents binary message 1, then the Gaussian's distribution mean is 5)

Now the chance for an error occurring is the probability that the voltage of the message falls on the wrong decision region (for example, message 0 with voltage 0, exceeds the threshold voltage of 0.5). The chance of the voltage exceeding the threshold voltage translates to simply finding the Gaussian distribution tail probability at the wrong region, which is where the need for using the Q-function comes. This concept is visualized in Fig. 42

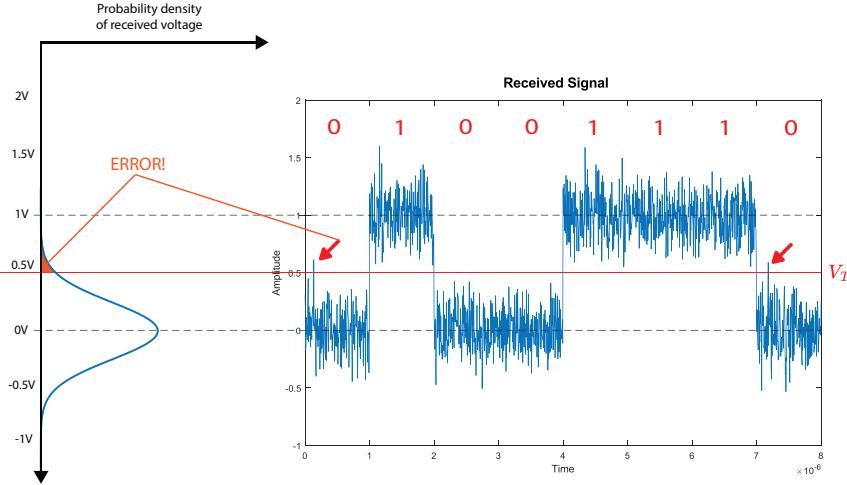


Figure 42: Illustration of the probability distribution of AWGN noise of message 0, showcasing that the Gaussian distribution tail of the noise which exceeds the decision region, represents the probability that an error will occur (hence the message falling into the wrong region)

Now if we also plot the Gaussian distribution of message '1' centered around 1V then we get the

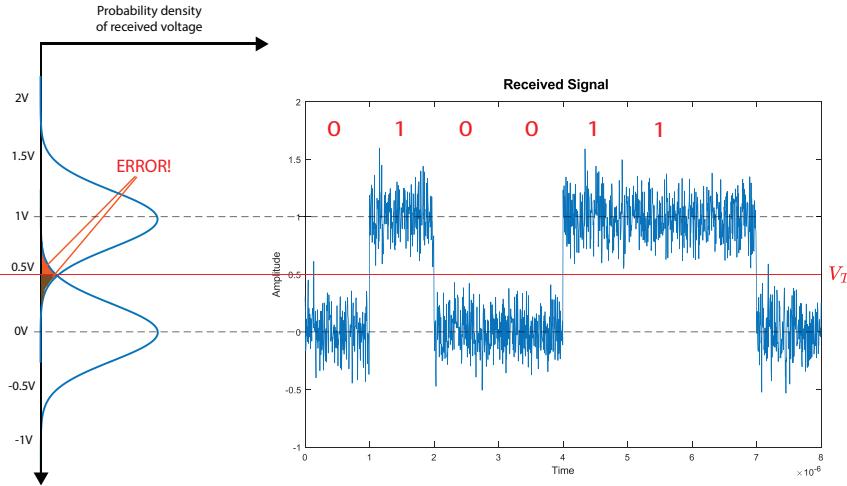


Figure 43: Illustration of the probability distribution of AWGN noise for both messages 0 and 1

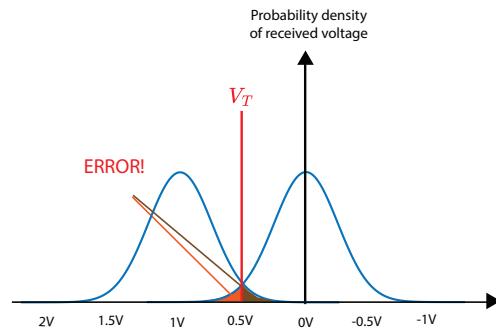


Figure 44: The probability distributions of the messages under AWGN noise

To find the probability of error  $P_e$ , we can formulate the problem as

Probability of error =

$$(\text{Probability of message 0}) \cdot (\text{Probability that message 0 falls in the wrong region (so higher than } V_T)) \\ + (\text{Probability of message 1}) \cdot (\text{Probability that message 1 falls in the wrong region (so lower than } V_T)) \quad (81)$$

Which mathematically can be described as

$$P_e = \Pr(0) \cdot P(x > V_T | 0) + \Pr(1) \cdot \Pr(x < V_T | 1) \quad (82)$$

Supposing that we are equally likely to receive a 0 or 1, and by using the Q-function we can write  $P_e$  as

$$P_e = \frac{1}{2}Q\left(\frac{V_T - V_0}{\sigma}\right) + \frac{1}{2}Q\left(\frac{V_1 - V_T}{\sigma}\right) \quad (83)$$

which finally can be expressed as (*derivation out of the scope of the course*):

$$P_e = Q\left(\sqrt{\left(\frac{S}{N}\right)_{in}}\right) \quad (84)$$

Equation (84) relates the probability of error with the SNR at the input of the receiver.

### 5.5.7 Relating $SNR_{in}$ and $SNR_{out}$

The  $(SNR)_{in}$  is the SNR of the received signal at the input of the receiver stage, which mainly deals with the channel noise (AWGN)

Furthermore, the  $SNR_{out}$  is the SNR at the output of the receiver, which is dependent on how the initial signal was quantized (via the term  $M^2$ ) but also how the digital signal was received (via the term  $P_e$ ).

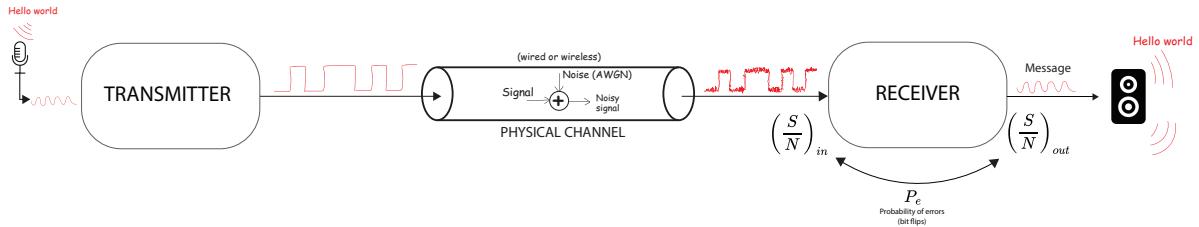


Figure 45: Simplistic view of transmitter-receiver chain, showing how probability of error relates  $SNR_{in}$  and  $SNR_{out}$

It may seem at first that by increasing the number of bits per sample  $n$ , the output SNR increases, however by adding more bits per sample, we may note that the  $SNR_{in}$  decreases because the bandwidth increases, meaning more noise power will also be present at the input. This effect can be noticed in Fig. 46.

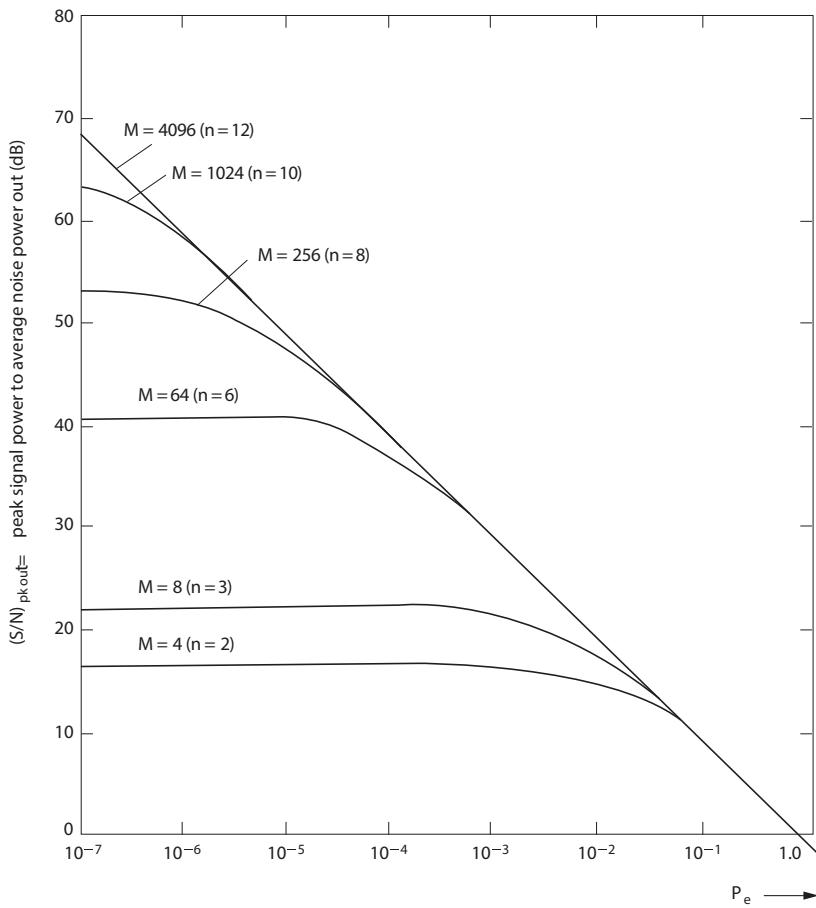


Figure 46:  $SNR_{out}$  of a PCM system as a function of  $P_e$  and the number of quantizer steps  $M$

### Exercise 3: Computing required transmit power

Suppose we have a digital communication system, for which an audio signal with a bandwidth of 20 kHz is being transmitted. The signal is quantized on the transmitter side using an A/D converter of 10 bits per level ( $n = 10$ ), and sampled at a sampling frequency  $f_s = 44.1 \text{ kHz}$ .

After some experimentation, it was found that the channel has a noise spectral density of  $\frac{N_o}{2} = 10^{-8} \text{ W/Hz}$ . Furthermore, assume spectral efficiency  $\eta = 1$ .

At the receiver, it is required that  $SNR_{out} = 30 \text{ dB}$ . Compute the transmit power of the transmitter antenna to be used to meet the system requirements.

**Solution:** In order to find the required transmit power to achieve the required  $SNR_{out}$ , we need to start from the end of the communication chain and work backwards. That is, we first find minimum  $P_e$  needed, and then using  $P_e$  we can compute a  $SNR_{in}$  value. Once we have a  $SNR_{in}$  value, we then need to find the noise power in the system using given noise spectral density, and from there we can then finally compute the required transmit power.

- 1) **Computing minimum probability of error** - We start with Eq. 67, and solve for  $P_e$

$$\left(\frac{S}{N}\right)_{out} = \frac{M^2}{1 + 4P_e(M^2 - 1)}$$

Solving for  $P_e$  leads to:

$$P_e = \frac{M^2 - \left(\frac{S}{N}\right)_{out}}{4 \left(\frac{S}{N}\right)_{out} (M^2 - 1)}$$

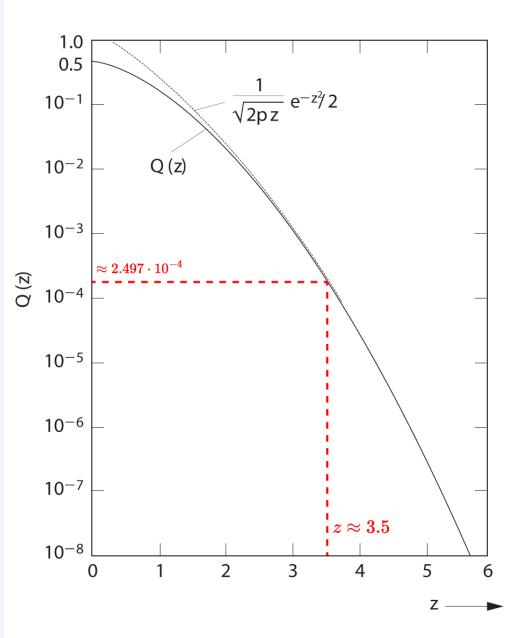
Now by plugging in the numbers,  $M = 2^n = 2^{10} = 1024$ . We also need to convert  $\left(\frac{S}{N}\right)_{out}$  from dB scale to power scale, which is done by  $10^{\frac{30}{10}} = 1000$ . Finally we have

$$P_e = \frac{(2^{10})^2 - 1000}{4 \cdot 1000((2^{10})^2 - 1)} = 2.497 \cdot 10^{-4}$$

- 2) **Finding required  $SNR_{in}$**  - such that it yields a  $P_e$  of at least  $2.497 \cdot 10^{-4}$ . From subsections 5.6 and 5.7 of the Digitization chapter, we can see that  $SNR_{out}$  is related using the Q-function to  $P_e$  as given in Eq. 84

$$P_e = Q\left(\sqrt{\left(\frac{S}{N}\right)_{in}}\right)$$

But how do we compute the Q-function? We can use the Q-function plot of Fig. 40 to find what approximate  $z$  value gives the  $P_e$  of  $2.497 \cdot 10^{-4}$ .



We can see from the figure above that  $z = 3.5$  will approximately yield our required  $P_e \approx 2.497 \cdot 10^{-4}$ . Now to find a  $\left(\frac{S}{N}\right)_{in}$  value:

$$P_e = Q(z) \equiv Q\left(\sqrt{\left(\frac{S}{N}\right)_{in}}\right)$$

$$z = \sqrt{\left(\frac{S}{N}\right)_{in}}$$

$$\left(\frac{S}{N}\right)_{in} = 12.25$$

So we find that we need a value of  $SNR_{in}$  of at least 12.25 (or 10.88 dB), in order to satisfy our  $P_e$  requirements, for which it directly satisfies our  $SNR_{out}$  requirement.

- **3) Computing required transmit power** Now that we have a value for the required  $SNR_{in}$  at the receiver we can compute what our  $S$  (transmit signal power) has to be to reach the required  $SNR_{in}$  at the receiver

However, we need to find first the noise power level in the channel. We are given the noise spectral density  $\frac{N_o}{2}$  and by recalling equation 78 we have that

$$N = \left(\frac{N_o}{2}\right)B_{pcm}$$

By using Eq.65 we can find the bandwidth of our digital signal  $B_{pcm}$

$$N = \left(\frac{N_o}{2}\right)(n \cdot f_s \cdot \frac{1}{\eta})$$

$$N = (10^{-8})(10 \cdot 44.1 \cdot 10^3) = 0.0044 \text{ W}$$

Finally, we can find the transmit power required to satisfy our  $SNR_{out}$  conditions,

$$\frac{S}{N} = 12.25$$

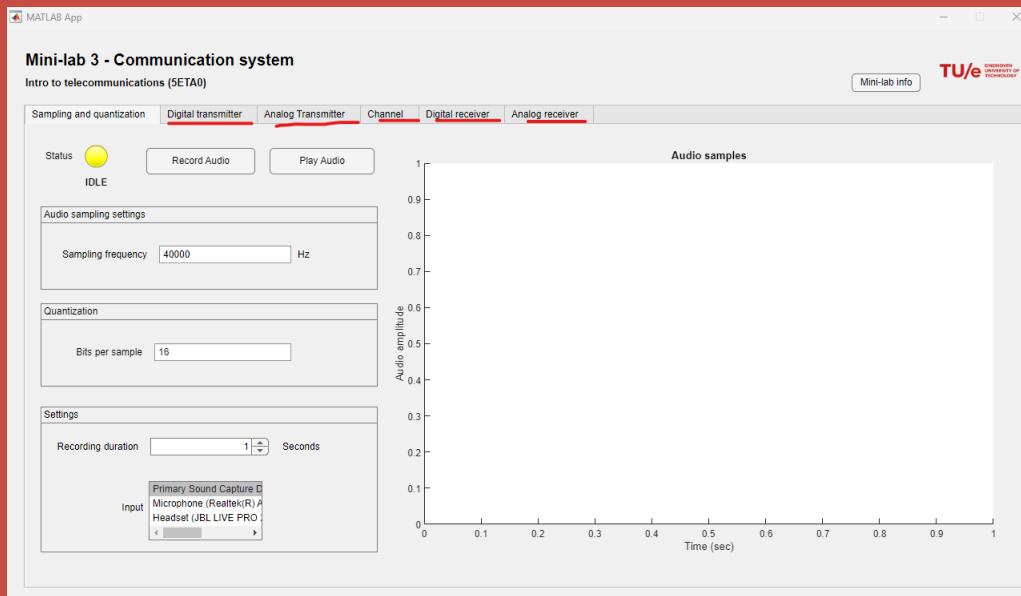
$$S = 12.25N = (12.25)(0.0044) = 0.0540 \text{ W}$$

So in order to have a  $SNR_{out} = 30 \text{ dB}$  we need to have a signal transmit power of 54 mW.

## Minilab exercise 5.2 - Digital communications under noise

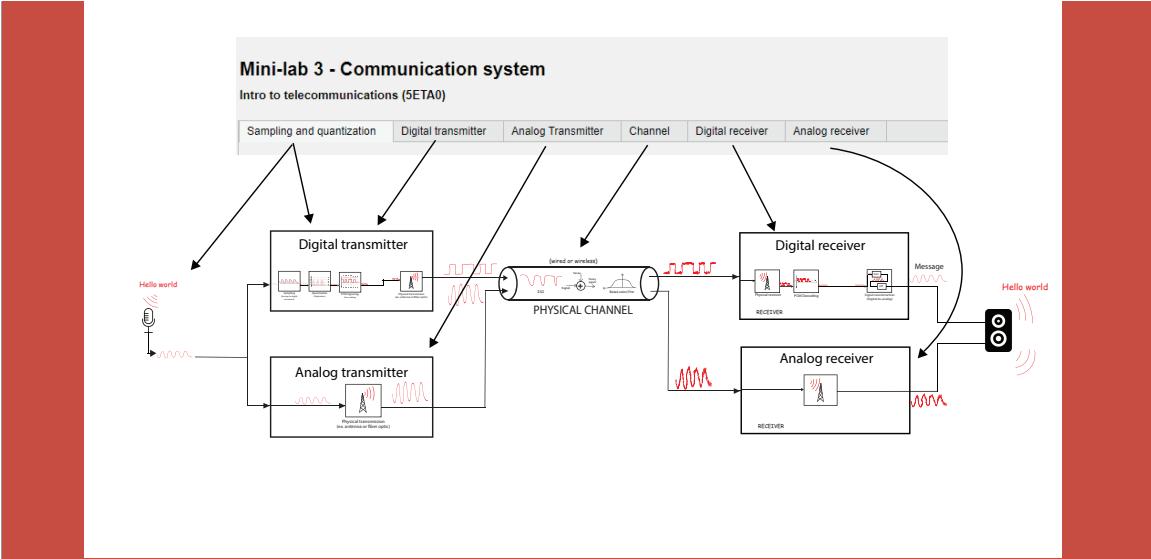
This mini-lab exercise requires you to use Mini-lab 3 (the communication system) on MATLAB.

When you open Minilab 3 you will be confronted with the following view:



It may at first seem similar to Minilab 2 however, you may see that there are more tabs than in Minilab 2, this is because in this minilab we consider the whole communication chain. We will first spend some time to understand how the minilab works.

The figure below shows the communication chain, and what each tab in the minilab represents in the chain. It additionally includes an analog transmitter and analog receiver, which both get transmitted through the same channel. The reason for including basic analog transmitters is that you can understand why digital communications are more robust in comparison to analog communications.



In the 'Sampling and Quantization' tab, you may change the settings, such as sampling frequency and the number of bits per sample, and it's also the place where you record the audio that you want to transmit. We highly advise keeping the recording duration to only 1 second, because when performing digital transmission and receiving, including a higher recording duration will lead to more data being processed, which might make the minilab run slower.

In the 'Digital transmitter' tab, you may choose the line coding (see Fig. 58 for all the line codings), and also the transmitter power in watts. You will also see details as to how much bandwidth each digital transmission is using based on  $f_s$  and  $n$ . You may also use the time knob to see the transmitted signal at a different time. You will also see the bitrate and symbol rate (this concept is introduced later)

In the 'Analog transmitter' tab you may only choose the transmitter power, and it will transmit the audio you recorded in the first tab.

In the 'Channel' tab you may change the noise spectral density  $N_o/2$  of the channel, which depicts how noisy is the channel.

In the 'Digital Receiver' tab you may see how the digital signal was received and how it was decoded. You can also see the probability of error  $SNR_{in}$  and also  $SNR_{out}$ . In this tab, you may also play to hear the decoded digital audio

In the 'Analog Receiver' tab you can see the received analog message and you may play the received analog audio.

Now we can start with the minilab exercise, we will first demonstrate why digital signaling is more robust than analog signaling and then we will proceed to show how adding more bits per sample will not improve the  $SNR_{out}$  but rather make it worse.

- 1) Open the 'Sampling and Quantization' tab, change the number of bits per sample to 8, and then record your voice. Make sure to keep the recording duration to just 1 second.

- 2) Once you have recorded, open the 'Digital transmitter' tab. In this tab, change the line coding to 'Polar RZ' and set the transmitter power to 10 Watt. Then go to the 'Channel' tab and set the noise spectral density to  $3e - 07$ . Then go again back to the 'Digital transmitter' tab and click the transmit button.
- 3) After you do this click on the button 'Transmit data' to transmit the data. You will have to wait about 20 seconds (depending on your computer) until the waveform shows up.
- 4) After the waveform shows up, go to the 'Digital receiver' tab, and play the audio. Take a note of the  $SNR_{out}$ . What do you hear? Is the audio quality poor? If so, from where does this noise come from? (hint: Its not because of channel noise)
- 5) Now, go back to the 'Sampling and Quantization' tab, and change the number of bits per sample to 16. Then, move to the 'Digital transmitter' tab and click the transmit button. After the new waveform shows up, go to the 'Digital receiver' tab and play the sound. Take note on the  $SNR_{out}$ . Do you hear a better-quality audio? Did the  $SNR_{out}$  value improve? Can you explain why increasing bits per sample increases audio quality (hence  $SNR_{out}$ )?
- 6) Now repeat step 5), but change the number of bits per sample to 24. Do you hear better-quality audio than with 16 bits per sample? Did the  $SNR_{out}$  value improve or get worse? Can you explain what is happening with the  $SNR_{out}$  value?
- 7) Now repeat step 5) (so keep the bits per sample to 16 and re-transmit). Now open the 'Analog transmitter' tab, and click transmit with the same power (so 10 watts). Go to the digital receiver play the audio, then go to the analog receiver and play the audio. At which receiver does the audio sound better? Why? Can you explain why digital is more robust than analog?
- 8) Start again by recording your voice, make sure to speak something (set n=16). Now on the channel settings change the noise spectral density to  $1e-06$ . Re-transmit both the digital and analog waveforms. Go to the digital receiver, play the audio, then go to the analog receiver and play the audio. At which receiver does the audio sound better? Why? What would you choose in this case analog or digital transmission? Explain your choice.

## 5.6 Instruction Exercises — Digitization

The solutions to these exercises may be found under the page **5ETC0 Canvas Page Modules** → **Week 2** → **Pulse Code Modulation (PCM) digitalization**

### Problem 4 (*Video solution available*)

For a PCM system is required that the signal to noise ratio at the output of the receiver is 34 dB in case of a uniformly distributed input signal of the quantizer. The number of levels of the quantizer is  $M = 2^n$  where  $n$  is the number of bits per sample.

- What is the minimum value of  $n$  to meet the system requirement?
- What signal to noise ratio can be tolerated at the input of the receiver in that case?

### Problem 5 (*Video solution available*)

A video signal with a bandwidth of 5 MHz is transmitted over a baseband channel having the same bandwidth, with white Gaussian noise added to the video signal. The signal-to-noise ratio of the received analog signal is 20 dB.

To improve the transmission quality the video signal is digitized to a PCM signal; it is sampled, quantized and coded in 400 levels. For the transfer, only the bandwidth of the channel is increased. The transmit power remains the same.

- What is the minimum bandwidth required for transmission of the PCM signal and with what pulse shape is this accomplished?
- Calculate in that case the bit error rate of the received digital signal.
- Calculate the signal-to-noise ratio of the recovered video signal at the output of the PCM receiver.

### Problem 6 (*Video solution available*)

A music signal with a spectrum till  $f_x = 16$  kHz is transmitted using PCM over a baseband channel with no bandwidth restrictions, but during the transfer white Gaussian noise is added. The requirement is that at the output of the PCM system, the signal-to-noise ratio  $(S/N)_O$  should have a minimum value of 50 dB in case of an uniform distributed input signal of the ADC.

- How large, the number of bits per sample  $n$  of the PCM system and the bandwidth  $B_T$  at the input of the digital receiver at least should be in order to meet the system requirement?
- If under these conditions, the exact requirement is met, how large is the signal-to-noise ratio  $(S/N)_R$  at the input of the receiver?
- Now the signal power is increased by 3 dB. How large is the signal-to-noise ratio at the output of the receiver now, if  $n$  is unmodified?
- What is the maximum achievable  $(S/N)_O$  with another choice of  $n$  and  $B_T$ ?

### **Problem 7** (*Video solution available*)

An analog video signal with a maximum frequency of 2 MHz is sent over a transmission line connection. The transmit power is 1 Watt. The transmission loss on the connection is 21 dB. The two sided spectral density of the received noise is  $\frac{N_o}{2} = 2 \times 10^{-11}$  Watt/Hz. The receiver is noise-free, and the receive filter is an ideal low-pass filter.

a) What is the signal-to-noise ratio at the input of the receiver?

The connection is now digitized, and the video signal is transmitted by means of a (binary) PCM system. The quantization is uniform, the number of bits per sample is 5 and the transmission system is using ideal sinc pulses for transmission. The transmission power remains the same (1 Watt), but the bandwidth of the receive filter is adapted, of course.

b) What is the SNR at the input the PCM receiver?

c) What is at the output of the PCM receiver the signal to noise ratio, SNR, of the recovered video signal?

You will now have the freedom to adjust the number of bits per sample (and the transmission bandwidth) to achieve optimal transmission.

d) What is the number of bits per sample for a maximum signal-to-noise ratio at the output of the PCM receiver and what is the signal-to-noise ratio?

The power density of the video signal is uniform over the whole band, however the amplitude of the video signal is not uniformly distributed: the signal has a peak-to-average ratio of 10 dB.

e) How can you use this to increase the quality of the transmission system even further?

### **Problem 8** (*Video solution available*)

In a digital transmission system an analog music signal is sampled at a frequency of 16 kHz, the samples are binary coded and transmitted as a PCM signal. The number of levels of the A/D converter  $M = 64$ . The input signal of the ADC is uniform distributed.

a) What is the maximum achievable signal to noise ratio of the regenerated signal at the output of the receiver?

The transmission channel has an effective bandwidth of 75 kHz, and the received power is 8 mW. By the transmission channel and the receiver, white Gaussian noise with a power spectral density of  $1.4 \times 10^{-8}$  W/Hz is added.

b) Calculate successively: 1. The signal to noise ratio at the input of the receiver, 2. The bit error rate of the detector and 3. The resulting signal to noise ratio at the output of the detector.

Now, the number of levels of the A/D converter is decreased to 32 without adapting the signal power and the bandwidth of the transmission system.

c) What is the resulting signal to noise ratio at the output of the receiver?

Now, the number of levels of the A/D converter is increased to 128, without adjusting the transmission parameters.

d) What is the effect on the signal quality in your opinion? Explain your answer.

## 6 Digital signaling

### 6.1 Learning objectives

Students completing this chapter should have learned:

1. Can calculate bit rate, baudrate and bandwidth for any given set of initial parameters.
2. Understand the difference between bits per sample and bits per symbol. Understand the limitations of multi-level modulation.

### 6.2 Motivation

In the previous section (digitization) we saw how an analog signal is converted into a digital binary stream (PCM). We also analyzed the bandwidth of PCM signals, and we saw that there is a dependence of the PCM signals based on the digital waveform we use (such as Polar NRZ, Unipolar NRZ etc.). In this section, we will see a vectorial representation of digital signaling, the concepts of multilevel signaling (which is used in every practical digital communications system), and the concept of symbol rate.

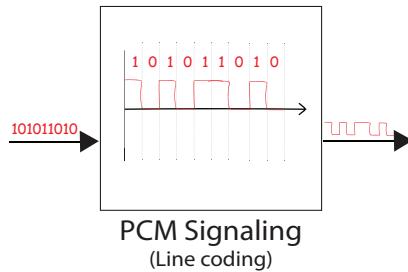


Figure 47: Digital signaling and line coding - Topic map location

### 6.3 Multi-level signaling

Now, to further understand the fundamentals of digital data transmission, we must understand how we prepare and transmit this information. These are the steps taken we have covered so far:

- Via our ADC, we converted the analog signal (e.g: audio wave) to a sequence of bits.
- Depending on the resolution (number of quantization bits  $n$  and the sampling frequency of our ADC, we may have more or less information which needs to be transmitted (so a smaller or higher Bit Rate R).

As a result we have a bit stream which is ready to be transmitted. One option would be simply sending the bits "one by one". So, for example, we could think of a digital "1" as being a pulse of +1V, and a digital "0" as simply -1V. This is commonly known as Polar NRZ transmission (see Fig. 58).

However, we could also choose to send two bits each time we transmit. For example, the sequence of bits "00" could correspond to -1V, "01" to -0.33V, "10" to +0.33V and "11" to +1V. Hence, we define multiple levels for the circuit output. Since we are starting with binary information, bunching bits together into a more complex bit sequence of 2 or more bits results in a more efficient use of the channel. This concept is illustrated in Fig. 48 where we can see that keeping the time between changes of the output waveform fixed results in halving the time needed to transmit the information. Since now every change of the waveform does no longer represent a single bit we call each new voltage level a symbol

and measure the amount of symbols with the unit symbol rate.

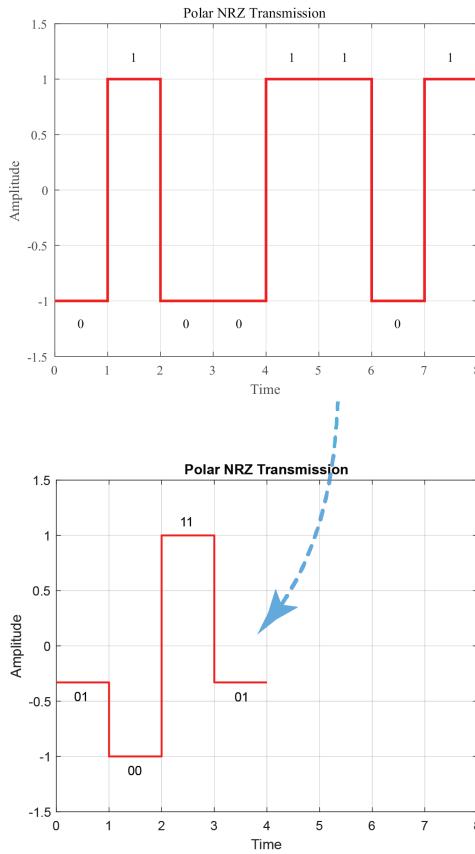


Figure 48: Figure illustrating the conversion of a digital transmission from 2 levels to 4 levels, showing that you can transmit the same amount of information in a faster time.

We define  $l$  as **the number of bits per level**, meaning the number of bits each level represents, and  $L$  as **the number of levels**.  $l$  and  $L$  are related as

$$L = 2^l \quad \text{or} \quad l = \log_2(L) \quad (85)$$

Moreover, with multilevel signaling, less bandwidth is needed for the same information transfer capacity (is symbol rate remains unchanged) or more information can be sent at the same time, if we increase the number of bits each symbol carries.

### 6.3.1 Advantages and Disadvantages of multilevel signaling

Using multiple bits per level can increase our data transmission rate. It would seem reasonable that increasing the number of levels to a very high number is the next logical step.

However, we must be aware of channel effects, and how this will affect what we obtain at the other end of our communication link, so at the input of the receiver.

For example we can simulate the effects of noise, on both the case where  $l = 1$  (so "one bit per symbol" transmission), and the case where  $l = 2$ , so we transmit two bits in each symbol.

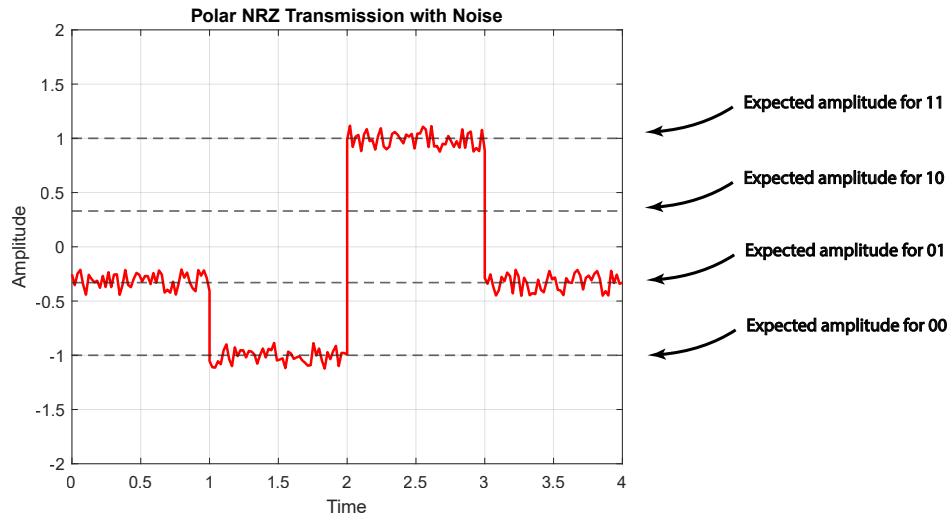


Figure 49: 4-level (2-bit) polar NRZ transmission under low noise—distinct amplitude levels allow accurate decoding.

As we can see in Figure above, there are 4 possible levels the amplitude can take (corresponding to all possible combinations of two bits).

We see that when the noise magnitude is small (minor variations in the waveform), we can still easily tell which combination of bits was sent. This is because the amplitude of the incoming waveform, whilst time-variant, stays relatively close to the expected level.

In figure, the effect of noise is stronger. In fact it is sufficiently harsh that at the receiver, the bit sequence decoded will most likely not coincide with what was actually sent.

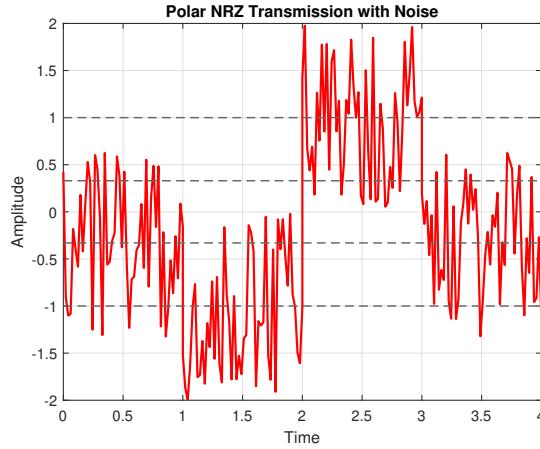


Figure 50: 4-level polar NRZ transmission under high noise—amplitude distortions lead to decoding errors.

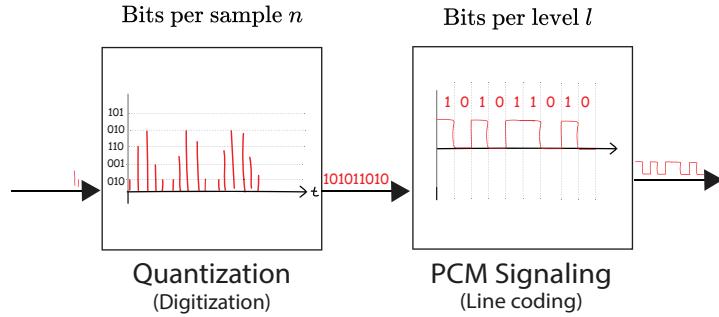
Now if we instead go back  $l = 1$ , so one bit per symbol, we see that there are only  $2^l = 2$  levels defined, one for +1V, and one for -1V, corresponding to digital 1 and 0. Applying the same situation of extreme noise, it's clear that even though the incoming signal is far from clean, we can discern accurately if either a 1 or 0 was sent.

### 6.3.2 Bits per sample and bits per level, the difference

Bits per sample  $n$  refers to the number of bits used to represent each sample of an analog signal in a digital domain. Bits per sample is a parameter in the digitization stage and determines the amount of information of the sampled analog signals.

On the other hand, bits per level  $l$  refer to the number of bits a level represents in multilevel signaling schemes. Bits per level is a parameter in the line coding stage.

Thus, bits per sample and bits per level are two different concepts in two different stages of the communication chain



## 6.4 Baud (or Symbol) rate

Symbol rate  $D$  refers to the number of signal changes (or symbols) per second in a communication channel. Each symbol in a transmission represents a specific number of bits  $l$ , and the baud determines how many symbols can be transmitted per unit of time. Symbol rate is equal to

$$D = \frac{N}{T_0} \quad (86)$$

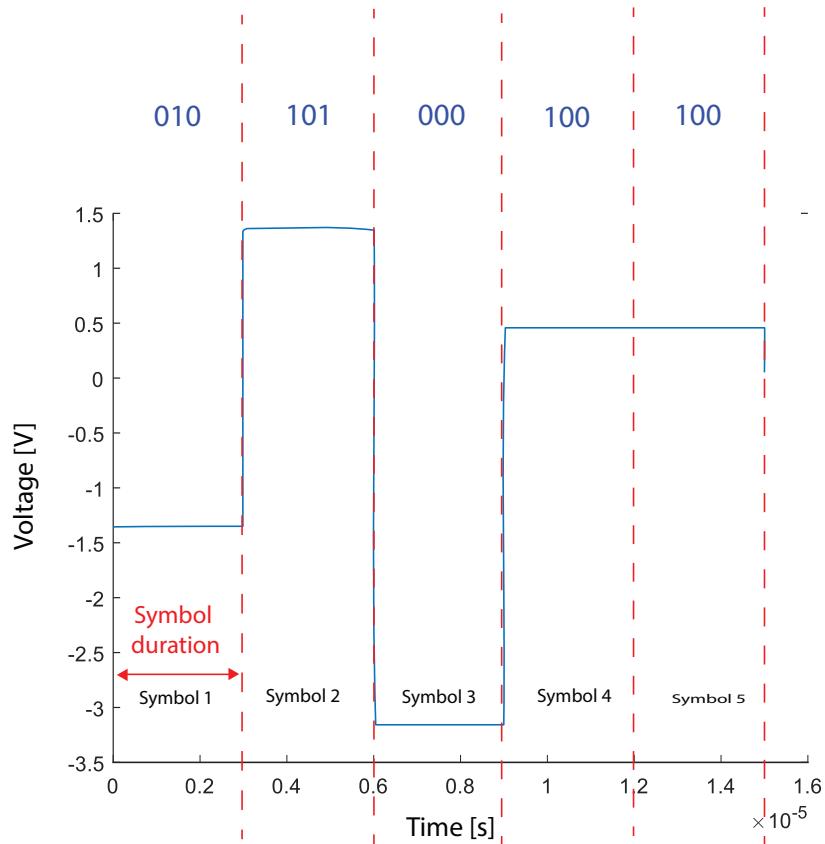
where  $N$  is the number of dimensions and  $T_0$  is the allocated time of a dimension.

It is important to note that the symbol rate is not necessarily equal to the number of bits per second  $R$ , as one symbol can represent multiple bits, as shown in multilevel signaling (see Fig. 51 for visualization). Thus, the relationship between them can be denoted as

$$R = lD = D \log_2(L) = \frac{N}{T_0} \log_2(L) = \frac{n}{T_0} \text{ [bits/s]} \quad (87)$$

e.g. for binary signaling (so where we only have two levels to represent 0 or 1),  $L = 2$ , hence  $l = \log_2(2) = 1$ , which means  $R = D$ .

Total Bits = 15, transmitted in 15  $\mu s$ , hence the bit rate is  $R = 1 Mbps/s$



Total Symbols = 5, transmitted in 15  $\mu s$ , hence the baud rate is  $D = 0.33 Msymbols/s$

Figure 51: Multilevel transmission with  $l = 3$ . The figure illustrates the difference between symbol rate and bit rate

#### 6.4.1 Bandwidth estimation

We recall the dimensionality theorem from sampling theory, which states that the number of orthogonal dimensions (independent pieces of information) to describe a signal with bandwidth  $B$  and time period  $T_0$  is:  $N_D = 2BT_0$ .

The relation between bandwidth  $B$  (in Hz) and symbol rate  $D=N/T_0$  (in symbols/s) is therefore (with  $N \leq N_D$  number of dimensions) :

$$D = \frac{N}{T_0} \leq \frac{N_D}{T_0} = 2B \quad (88)$$

so,

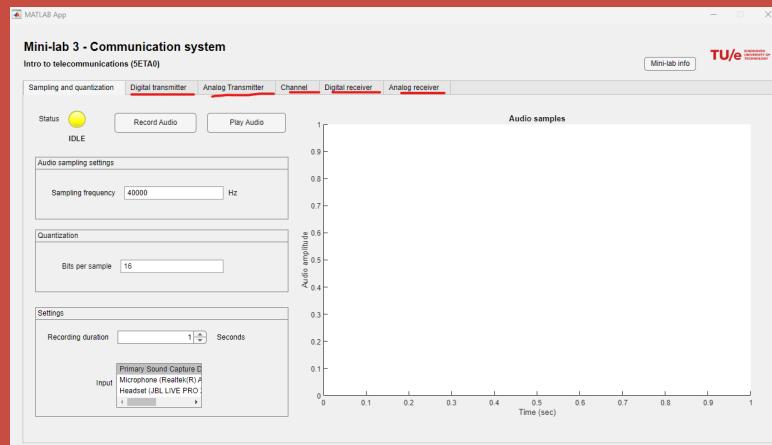
$$B \geq \frac{1}{2}D \text{ [Hz]} \quad (89)$$

This implies that the minimum bandwidth is half the symbol rate, and this minimum bandwidth is achieved when using sinc pulses, where the minimum spacing between pulses is the time between zeros (so  $1/2B$ ).

#### Minilab exercise 6.1 - Multi-level signaling

**This mini-lab exercise requires you to use Mini-lab 3 (the communication system) on MATLAB.**

**When you open Minilab 3 you will be confronted with the following view:**



In this minilab exercise, multilevel signaling will be demonstrated.

- 1) Start by recording your voice for a second, with configurations of  $f_s = 40000$  Hz and  $n=16$ . Go to the digital transmitter section, and choose 'Multilevel' signaling. Leave every other setting as default, and transmit the data.
- 2) Go to the digital receiver, and play the audio, does the audio sound correct? What do you observe with the value of  $B_{pcm}$  in comparison to other line codes?
- 3) Now on the digital transmitter tab, increase the number of bits per level to 5, and retransmit. Go to the digital receiver, and play the audio, does the audio sound correct? If the audio does not sound correct, can you explain what is happening?

## 7 Linecodes and their spectras

### 7.1 Learning objectives

Students completing this chapter should have learned:

1. Understand the concept of Power Spectral Density and its connection to the choice of line coding.
2. Able to quantitatively discuss the properties of a PSD for any of the 5 line codes specifically detailed in this chapter.
3. Able to calculate the impact of a choice of line code on the spectral efficiency and the required bandwidth of a communication channel.

### 7.2 Motivation

Line coding plays a crucial role in converting digital bits into waveforms for physical transmission. Building on the concepts introduced in the digital signaling chapter, where binary data was processed and transmitted, this chapter explains how different coding methods shape these signals to suit real-world channels. The chapter explores various techniques and discusses their impact on bandwidth usage and noise resistance.

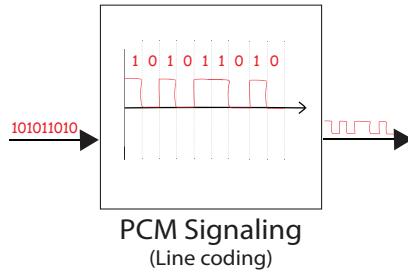


Figure 52: Digital signaling and line coding - Topic map location

### 7.3 Power spectral density (PSD)

Power spectral density (PSD) is a tool that shows how the power of a signal is spread over different frequencies. In many cases, the exact shape of the signal is not known in advance because it depends on the information being transmitted. Instead of using a simple Fourier transform on a single, fixed waveform, PSD is used to estimate the average power distribution over frequency for random or varying signals. This is very useful for estimating the bandwidth required for a transmission channel and its components.

#### 7.3.1 Derivation of PSD

A central idea in deriving the PSD is **autocorrelation**. Autocorrelation measures how similar a signal is to a delayed version of itself. In simple terms, it shows how the value of the signal at one time relates to its value at another time, where the delay is denoted by  $\tau$ . If a signal shows high similarity at a certain delay, this indicates a repeating pattern or periodicity.

For a continuous signal, the autocorrelation function  $R_w(\tau)$  is calculated by multiplying the signal with a shifted copy of itself and then averaging over time. The Fourier transform of this autocorrelation function gives the PSD, as expressed by:

$$F[R_w(\tau)] = \lim_{T \rightarrow \infty} \frac{1}{T} |W(f)|^2 = P_w(f) \quad (90)$$

where  $W(f)$  is the Fourier transform of the signal, and  $P_w(f)$  represents its power spectral density (see Appendix C for derivation of Eq.(90))

The Fourier transform converts the time-domain signal into its frequency components. When applied to the autocorrelation function, it shows how much power is present at each frequency, which is crucial for understanding the signal's overall bandwidth usage.

For digital signals that consist of discrete symbols, a discrete version of autocorrelation is used:

$$R(k) = \sum_{i=1}^I (a_n \cdot a_{n+k})_i P_i \quad (91)$$

Here,  $P_i$  is the probability of the  $i$ -th product  $(a_n \cdot a_{n+k})_i$ , and  $I$  is the number of possible pairs  $(a_n, a_{n+k})$ . This version takes into account the randomness inherent in digital signals.

Using the relationship between the Fourier transform of the autocorrelation function and the PSD (see (90)), the PSD for a random digital signal can be written as:

$$P_s(f) = \lim_{T \rightarrow \infty} \frac{|S_T(f)|^2}{T} = \frac{|F(f)|^2}{T_s} \sum_{k=-\infty}^{\infty} R(k) e^{j2\pi kfT_s} \quad (92)$$

In this equation,  $T_s$  is the symbol period, and the sum over  $k$  accounts for the contributions from all time delays.

After converting binary information from the quantizer and PCM stages into symbols, these symbols are transformed into physical waveforms for transmission. This process, known as **line coding**, uses various methods (such as Unipolar NRZ, Polar NRZ, Unipolar RZ, Bipolar RZ, and Manchester), each with its own PSD characteristics. These characteristics influence the system's performance, particularly in terms of bandwidth efficiency and noise tolerance.

## 7.4 Unipolar NRZ

Below you may see a visualization of a Unipolar NRZ transmission. Since it is a unipolar signaling waveform, only one power supply is needed. This is because either we use the supply voltage, or we use 0 (ground reference). Furthermore, following the analysis of the PSD, a DC component is present.

Its PSD is given as:

$$P_{\text{unipolar NRZ}}(f) = \frac{A^2 T_b}{4} \operatorname{sinc}^2(f T_b) [1 + \frac{1}{T_b} \delta(f)] \quad (93)$$

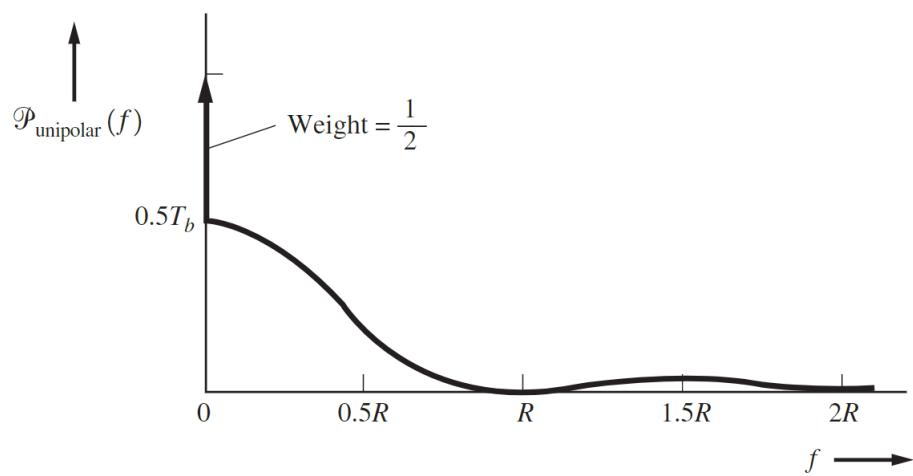
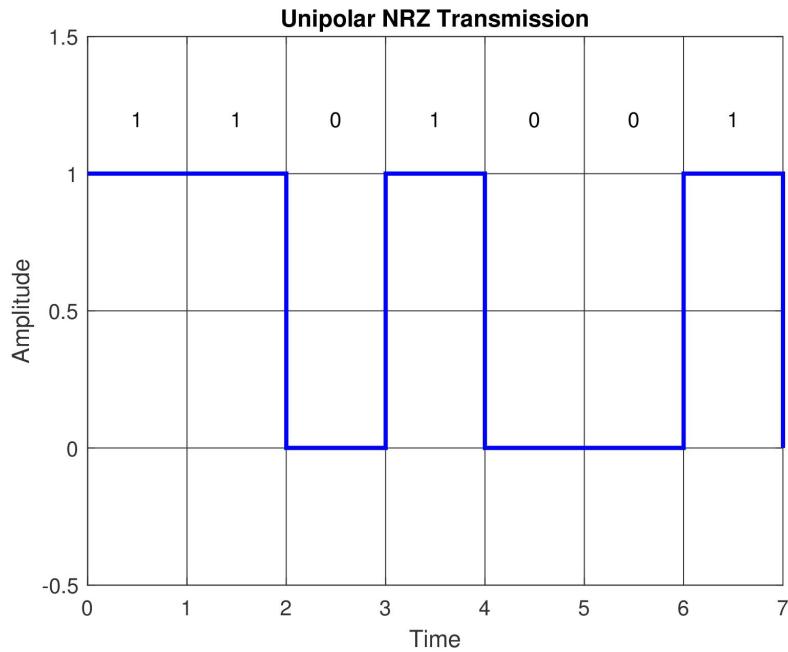


Figure 53: PSD of Unipolar NRZ

In order to calculate the eventual PSD we must first calculate the autocorrelation. The autocorrelation can be found for Unipolar NRZ in the following manner. We distinguish between two cases:

- The correlation with no displacement ( $k=0$ ).
- The correlation with a displacement of integer value ( $k \neq 0$ ).

For  $k=0$  there are only two possibilities (with equal probability) for the symbols (both are zero or both are one). For the case  $k \neq 0$  case we are looking at 4 possible permutations (with equal probability) for the case of two symbols with a distance  $k$ . For Unipolar NRZ we assume the  $a_n$  can either have an amplitude A or 0 Volts.

$$R(0) = \sum_{i=1}^2 (a_n a_n)_i P_i = \overbrace{A^2 \cdot \frac{1}{2}}^{11} + \overbrace{0 \cdot \frac{1}{2}}^{00} = \frac{1}{2} A^2$$

$$R(k) = \sum_{i=1}^4 (a_n a_{n+k})_i P_i = \overbrace{A^2 \cdot \frac{1}{4}}^{11} + \overbrace{0 \cdot \frac{1}{4}}^{01} + \overbrace{0 \cdot \frac{1}{4}}^{10} + \overbrace{0 \cdot \frac{1}{4}}^{00} = \frac{1}{4} A^2 \quad \text{for } k \neq 0$$

$$\sum_{k=-\infty}^{\infty} R(k) e^{j2\pi k f T_b} = \frac{1}{4} A^2 + \frac{1}{4} A^2 \sum_{k=-\infty}^{\infty} e^{j2\pi k f T_b} = \frac{A^2}{4} \left[ 1 + \sum_{k=-\infty}^{\infty} e^{j2\pi k f T_b} \right]$$

## 7.5 Polar NRZ

Below you may see a visualization of a Polar NRZ transmission. Since it is a polar signaling waveform, two power supplies is needed (positive and negative rails). Furthermore, following the analysis of the PSD, there is a large DC component.

$$P_{polarNRZ}(f) = A^2 T_b \operatorname{sinc}^2(f T_b) \quad (94)$$

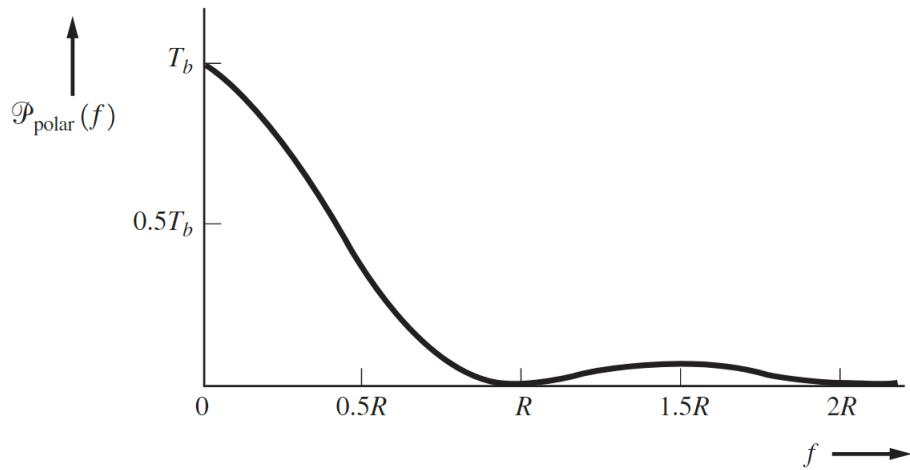
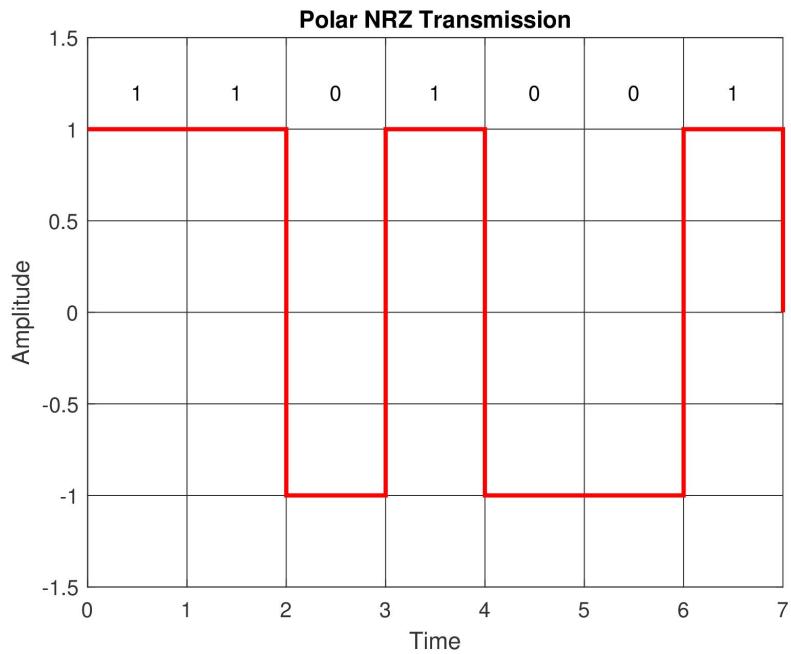


Figure 54: PSD of Polar NRZ

## 7.6 Unipolar RZ

Below you may see a visualization of a Unipolar NRZ transmission. Since it is a unipolar signaling waveform, one power supply is needed. Furthermore, following the analysis of the PSD, there is a DC component. The effects of 'Returning-to-Zero' are the doubling of bandwidth, as well as the presence of synchronization information.

$$P_{\text{unipolarRZ}}(f) = \frac{A^2 T_b}{16} \operatorname{sinc}^2\left(\frac{f T_b}{2}\right) \left[1 + \frac{1}{T_b} \sum_{n=-\infty}^{\infty} \delta(f - \frac{n}{T_b})\right] \quad (95)$$

Note that if we want to use variable duty cycle, the PSD is given as:

$$P_{\text{unipolarRZ}}(f) = \frac{A^2 d^2 T_b}{4} \operatorname{sinc}^2(f d T_b) \left[1 + \frac{1}{T_b} \sum_{n=-\infty}^{\infty} \delta(f - \frac{n}{T_b})\right] \quad (96)$$

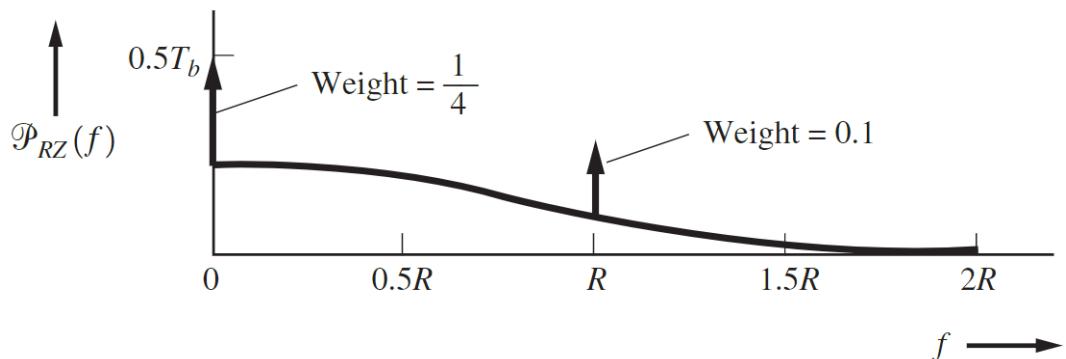
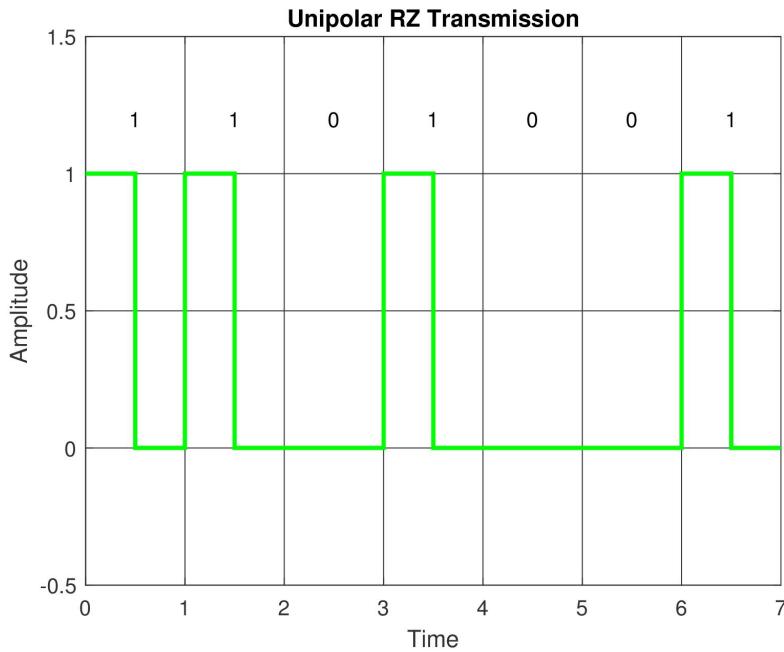


Figure 55: PSD of Unipolar RZ

## 7.7 Bipolar RZ

Below you may see a visualization of a Bipolar NRZ transmission. Since it is a bipolar signaling waveform, two power supplies are needed. Furthermore, following the analysis of the PSD, there is no DC component. Furthermore due to the 'Return-to-Zero' there are three possible voltage levels to be detected at the receiver (positive, 0, negative).

$$P_{\text{bipolar}_{RZ}}(f) = \frac{A^2 T_b}{4} \operatorname{sinc}^2\left(\frac{f T_b}{2}\right) \sin^2(\pi f T_b) \quad (97)$$

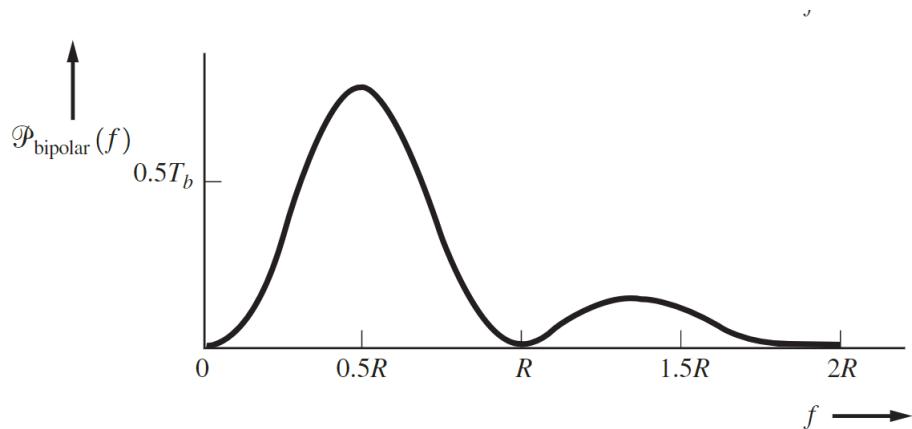
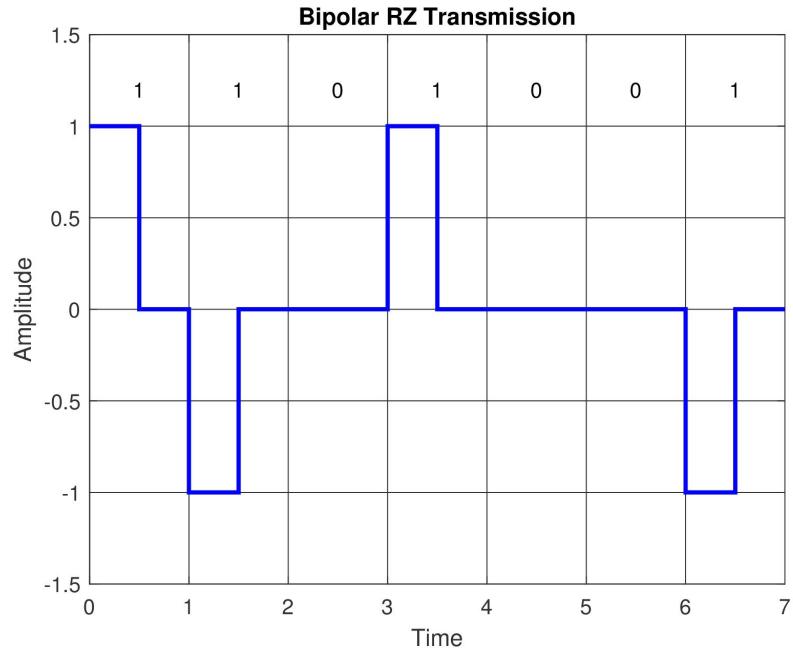


Figure 56: PSD of Bipolar RZ

## 7.8 Manchester

Lastly we can observe the Manchester linecode, with the special characteristic that a string of zero's will not cause loss of the clock signal. Furthermore it allows for error detection, since we can realize this whenever there are more than two incoming detected voltages (of transmitted bits) at the same level.

The PSD is given as:

$$P_{\text{Manchester}}(f) = A^2 T_b \operatorname{sinc}^2\left(\frac{fT_b}{2}\right) \sin^2\left(\frac{\pi f T_b}{2}\right) \quad (98)$$

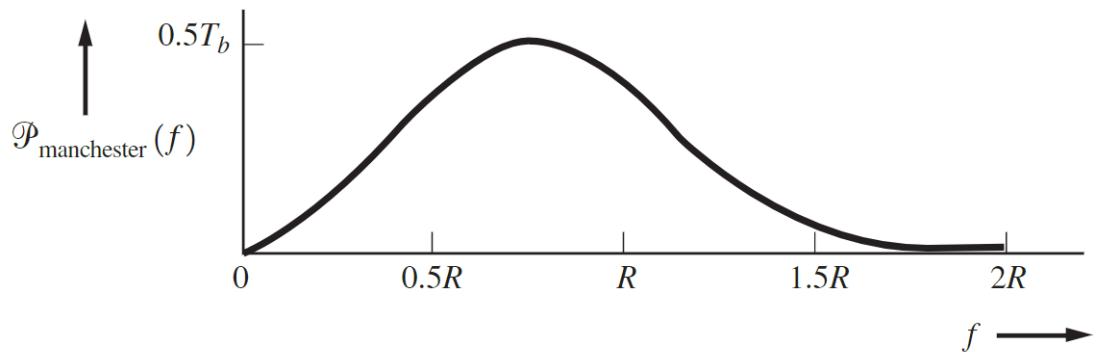
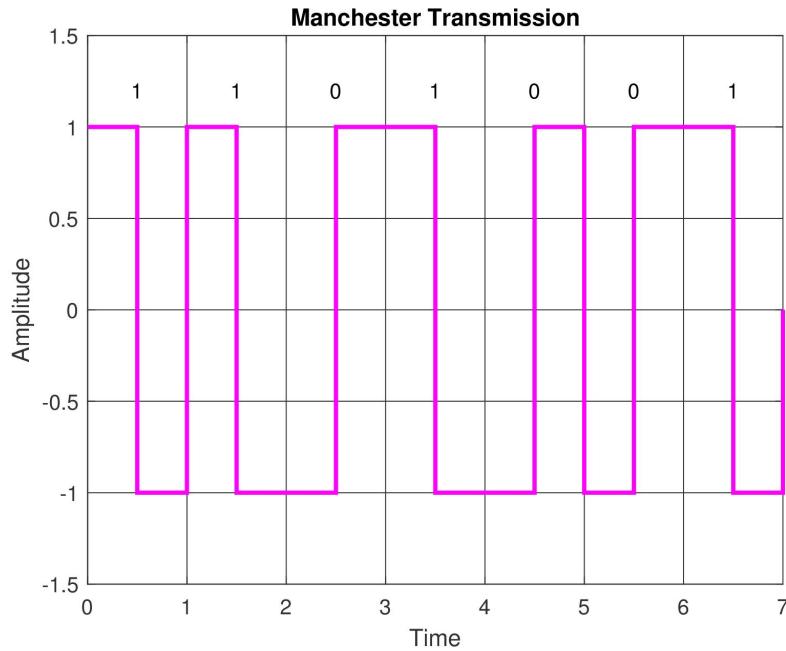


Figure 57: PSD of Manchester

## 7.9 Power spectra for multilevel signaling

Note that when moving towards  $l$  (bits per symbol) larger than unity (so any multilevel signal), we must also make the conversion for the PSD. This is additional to the conversion required for the spectral efficiency.

Hence we combine with the following:

$$D = \frac{1}{T_s} = \frac{1}{lT_b} = \frac{R}{l} \quad (99)$$

$$P_s(f) = \frac{\|F(f)\|^2}{T_s} \sum_{k=-\infty}^{\infty} R(k)e^{j2\pi kfT_s} \quad (100)$$

Obtaining thus that for a 8-level Polar NRZ signaling,  $l = 3$  and as such  $T_s = 3T_b$ .

$$\frac{|F(f)|^2}{T_s} = \frac{T_s^2 \text{sinc}^2(fT_s)}{T_s} = 3T_b \text{sinc}^2(3T_b f) \quad (101)$$

## 7.10 Spectral efficiency of linecodes

The table below gives the spectral efficiency for different lines codes assuming  $l=1$  or that  $D=R$ .

TABLE 3-6 SPECTRAL EFFICIENCIES OF LINE CODES

Code Type	First Null Bandwidth (Hz)	Spectral Efficiency $h = R/B$ [(bits/s)/Hz]
Unipolar NRZ	R	1
Polar NRZ	R	1
Unipolar RZ	2R	$\frac{1}{2}$
Bipolar RZ	R	1
Manchester NRZ	2R	$\frac{1}{2}$
Multilevel polar NRZ	$R/\ell$	$\ell$

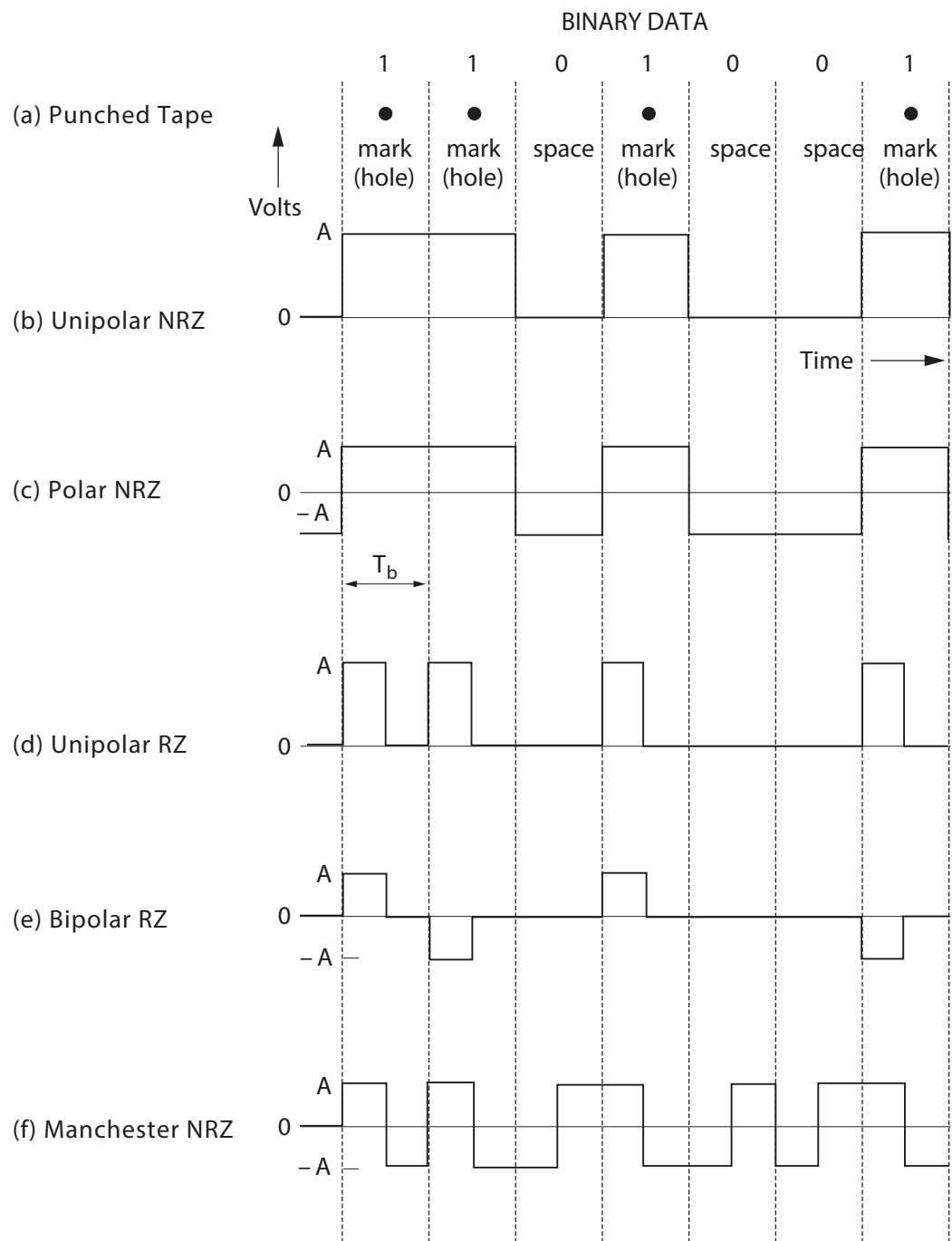


Figure 58: Summary of all line codes in this chapter

## 7.11 Instruction Exercises - Digital signaling and line codes

The solutions to these exercises may be found under the page **5ETC0 Canvas Page Modules** → **Week 2** → **E. Digital signaling and Line codes**

### **Problem 9** (*Video solution available*)

A transmission system has an ideal (rectangular) transfer function with a bandwidth of 4 kHz.

- a) How many symbols or numbers per second can be sent over this channel?
- b) ~~What is the theoretical maximum bit rate that can be transmitted with this channel? Motivate your answer. (Discard this subquestion)~~
- c) What is the maximum bit rate for the transfer if binary unipolar NRZ signals are applied?
- d) What is the power spectrum of a random sequence of these signals? Calculate the spectrum and give a sketch showing the characteristic points.
- e) We now replace the unipolar NRZ line coding by polar NRZ signals, while the transmitted power remains the same. Will this change the quality of the transfer? Explain your answer.

$$\sum_{k=-\infty}^{\infty} e^{j2\pi kfT_b} = \frac{1}{T_b} \sum_{k=-\infty}^{\infty} \delta(f - \frac{n}{T_b}) \quad (102)$$

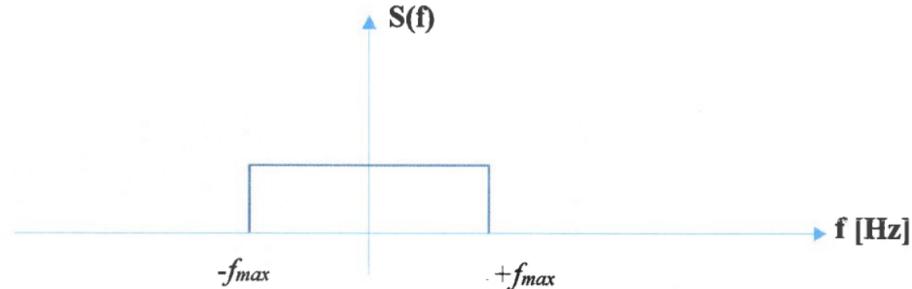
### **Problem 10** (*Video solution available*)

A music signal is converted to a PCM signal and sent through an AC coupled distortion-free transmission channel. The number of discrete levels of the quantizer is 4096. These levels are digitally coded and the information is transmitted using multi-level polar NRZ symbols. The number of levels per symbol is 16. The symbol rate of the coded signal is 132 kbaud.

- a) What is the sample rate of the sampler?
- b) What is the required transmission bandwidth?
- c) ~~What causes baseline wander in an AC coupled system? (Discard this subquestion)~~

**Problem 11** (*Video solution available*)

An analogue video signal is transported by means of a binary PCM transmission system. The line coding used is Unipolar RZ with 50% duty cycle. The video signal is sampled with a sampling frequency of 40 MHz. The samples are coded with 5 bits per sample. Assume delta sampling is used for sampling. The unsampled spectrum of the video signal is shown in the figure below:



Assume that  $f_{max}$  is 1/4 of the sampling rate!

- a) Draw the PCM data assuming line coding is Unipolar 50% RZ for the code words 01001 and 01111. Make sure to correctly identify the time constant needed in the sketch.

The PCM transmission system is designed to deliver a signal-to-noise ratio of 12 dB at the input of the receiver system. If the double-sided noise spectral density is given to be  $10^{-12}$  Watt/Hz and the transmitter sends out 1 Watt of signal power.

- b) What is the maximum attenuation that can be tolerated between the transmitter and the receiver so that the required SNR is achieved?  
c) What is the signal to noise at the output of the PCM receiver circuit?  
d) Would increasing the number of quantization levels improve the performance of the system?

## 8 Inter-symbol interference

### 8.1 Learning objectives

Students completing this chapter should have learned:

1. Understand that bandwidth limited channels will introduce signal distortions leading to inter symbol interference (ISI).
2. Can calculate the spectral efficiency and bandwidth requirement for using raised-cosine with a roll off factor  $r$ .
3. Understand why *Sinc* pulses offer the highest  $D/B$  ratio due to symbol overlap.

### 8.2 Motivation

In practical communication systems, the presence of 'Inter-Symbol Interference' (ISI) is an unavoidable phenomenon. ISI is inherently undesirable due to its potential to distort transmitted waveforms, introduce errors at the receiver, and impose limitations on transmission rates; therefore, its study is of paramount importance. Understanding and effectively managing this phenomenon becomes crucial for enhancing the reliability and efficiency of telecommunication systems.

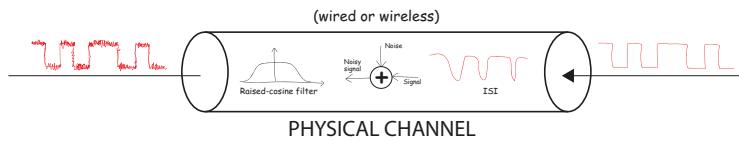


Figure 59: Topic map location

### 8.3 What is inter-symbol interference?

As we have seen previously, the absolute bandwidth of rectangular pulses is infinite. However, in every channel, there are bandwidth limitations, either imposed specifically (such as in wireless systems, where you can only use a certain amount of bandwidth) or the channel itself will act as a filter limiting the bandwidth. Filtering causes frequency components to be attenuated, which then in the time domain means the rectangular pulse will be less rectangular and more smoothed (hence spreading, supposing the filter is filtering high-frequency components). These bandwidth limitations will cause the pulses to spread in time, and the pulse for each symbol may be smeared into adjacent time slots, which can corrupt the information being carried in the pulse on the adjacent timeslot, by altering its actual amplitude and hence, cause ISI on the receiver. The concept of ISI is visualized in Fig. 60

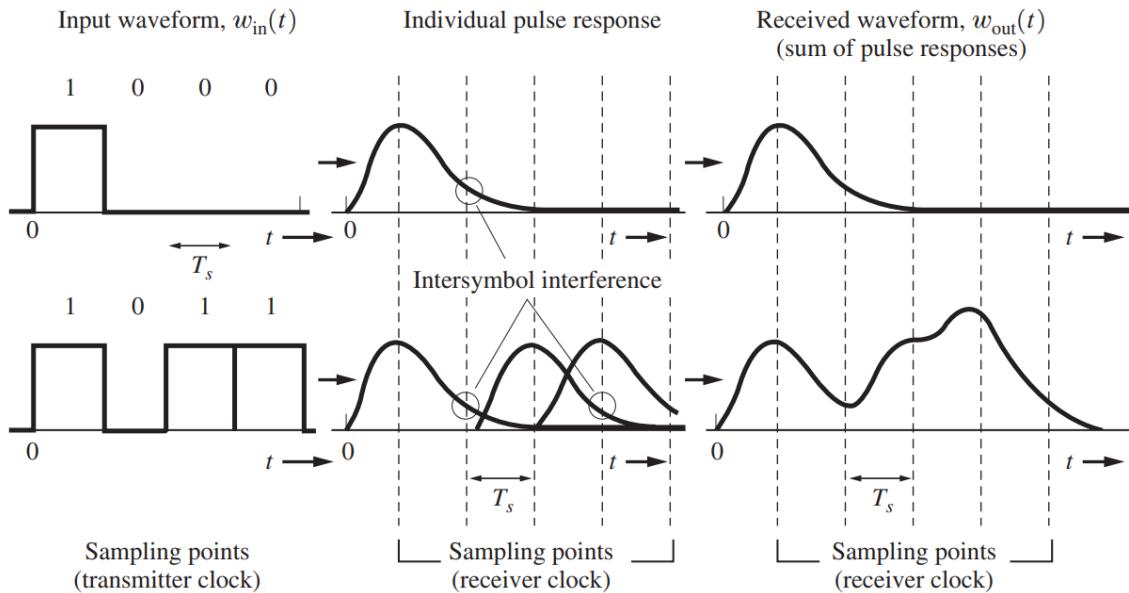
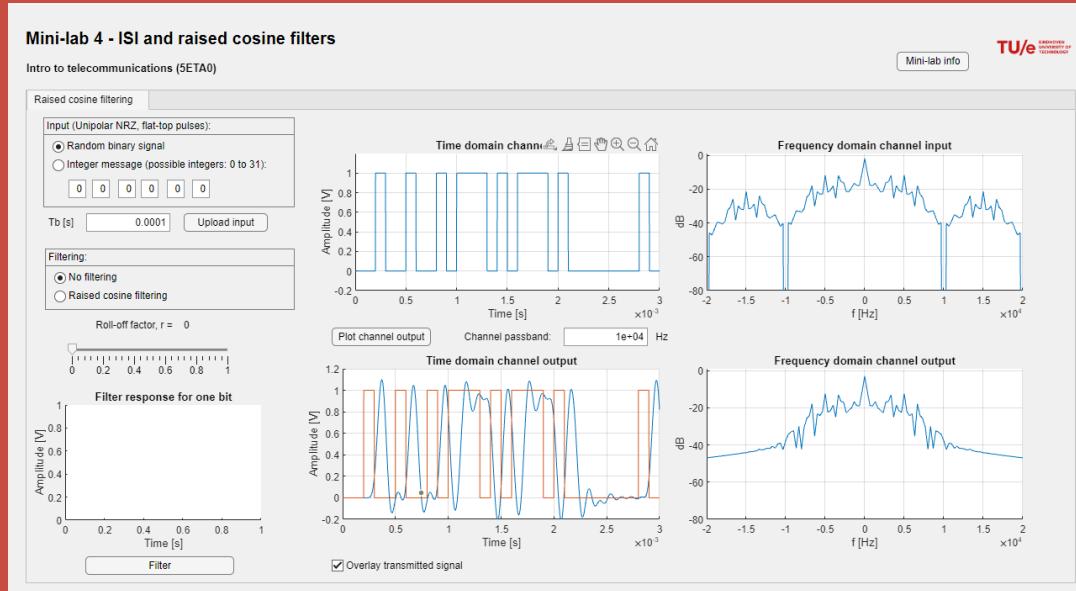


Figure 60: The concept of ISI visualized, for an individual rectangular pulse, and a stream of rectangular pulses. [2, ch. 3-6, p. 207]

## Minilab exercise 8.1 - ISI Visualization

**This mini-lab exercise requires you to use Mini-lab 4 - ISI (Intersymbol interference)**  
**When you open Minilab 4 you will be confronted with the following view:**



- 1) Leave the default settings and click on upload input to transmit the signal through the channel.
- 2) Click on 'Plot channel output' to see the received signal
- 3) Firstly, make sure to check the box '*overlay transmitted signal*', which helps you compare how the signal was transmitted and how it was received on the bottom-left plot. Now, decrease the channel passband frequency to 5000 Hz from 20000 Hz, in steps of 5000 Hz. Every time you decrease it, please click on '*plot channel output*' to see the effects. Can you notice intersymbol interference occurring?

## 8.4 Modelling ISI

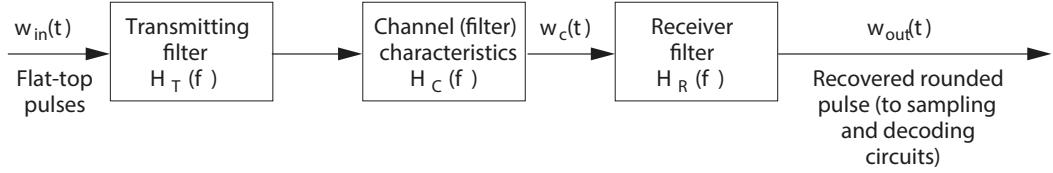


Figure 61: Base-band pulse-transmission system [2, ch. 3-6, p. 208]

Consider a digital signaling system (baseband) as in Fig. 61. Assume that the input pulses to this system  $w_{in}(t)$ , are as follows

$$w_{in}(t) = \sum_n a_n h(t - nT_s) = \sum_n a_n \delta(t - nT_s) * h(t) \quad (103)$$

where  $h(t) = \prod(\frac{t}{T_s})$  represent rectangular pulses. Now  $w_{out}$  can be written as the convolution of the input signal  $w_{in}(t)$  with the overall system transfer characteristic, which is the convolution between the transmitting filter  $H_T$ , channel filter characteristics  $H_C$  and receiver filter characteristics  $H_R$ .

$$w_{out}(t) = w_{in}(t) * h_T(t) * h_C(t) * h_R(t) = (\sum_n a_n \delta(t - nT_s)) * h_e(t) \quad (104)$$

where  $h_e(t)$  is the individual impulse response, written as

$$h_e(t) = h(t) * h_T(t) * h_C(t) * h_R(t) \quad (105)$$

And the representation in the frequency domain is

$$H_e(f) = H(f)H_T(f)H_C(f)H_R(f) \quad (106)$$

Essentially, Eq. 105, is the overall impulse response of the individual pulses that are transmitted with amplitude  $a_n$ , where the amplitude may take multi levels (in multi-level signaling). As we can see, if the input pulses  $h(t)$  are rectangular, then they go through channel filtering, which will cause some frequency components to be attenuated and hence smoothen the signal, which causes the signal to fall onto adjacent timeslots and cause ISI.

To have a  $h_e(t)$  which does not cause ISI on the sampling moments  $kT_s$ , it needs to be in the form of

$$h_e(\tau + kT_s) = \begin{cases} C & \text{for } k = 0 \\ 0 & \text{for } k \neq 0 \end{cases} \quad (107)$$

The equation above describes that every pulse should be only nonzero on its designated time-slot, and zero on every other adjacent time-slot. This is also known as a *zero-ISI equalization* or *Nyquist's first criterion*.

The output pulse shape is affected by the input pulse shape  $w_{in}(t)$ , the transmitter filter  $H_T$ , the channel filter  $H_C$ , and the receiving filter  $H_R$ . Because, in practice, the channel filter is already specified, the problem is to determine the transmitting filter and the receiving filter that will minimize the ISI on the rounded pulse at the output of the receiving filter.

## 8.5 Nyquists first method (Zero ISI)

The simplest waveform (or pulse) that satisfies Nyquist's first criterion is a *sinc pulse*. To see why a sinc pulse satisfies Nyquist's first criterion and does not spread, we recall Eqn. (108). The equation states that a transmitted pulse must only have a non-zero value on its designated timeslot for sampling, and zero on any other  $kT_s$  sampling timeslot. A ideal sinc pulse satisfies this condition. To demonstrate this concept, assume we want to transmit the letter 'N' (in binary 01001110) using sinc pulses. Figure 62 contains a plot of these sinc pulses transmitted every  $kT_s$

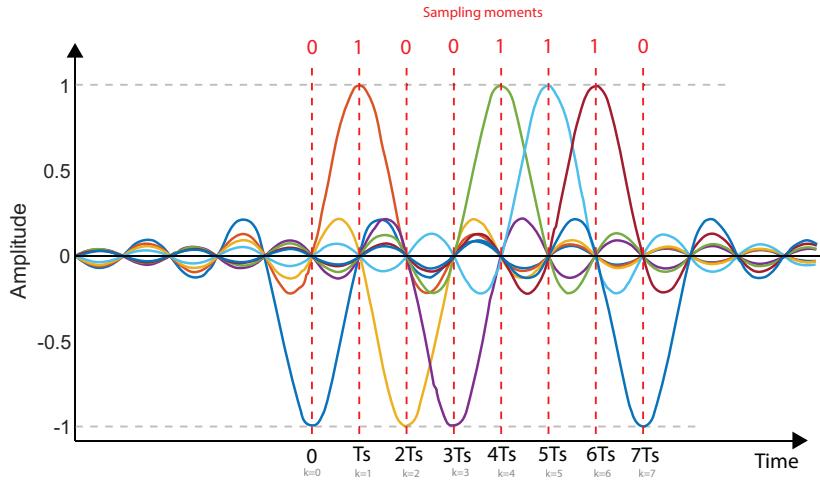


Figure 62: Sinc pulses representing letter N in binary, transmitted every  $kT_s$ .

Now, the final waveform (which is the sinc pulses summed together) is plotted in Fig. 63

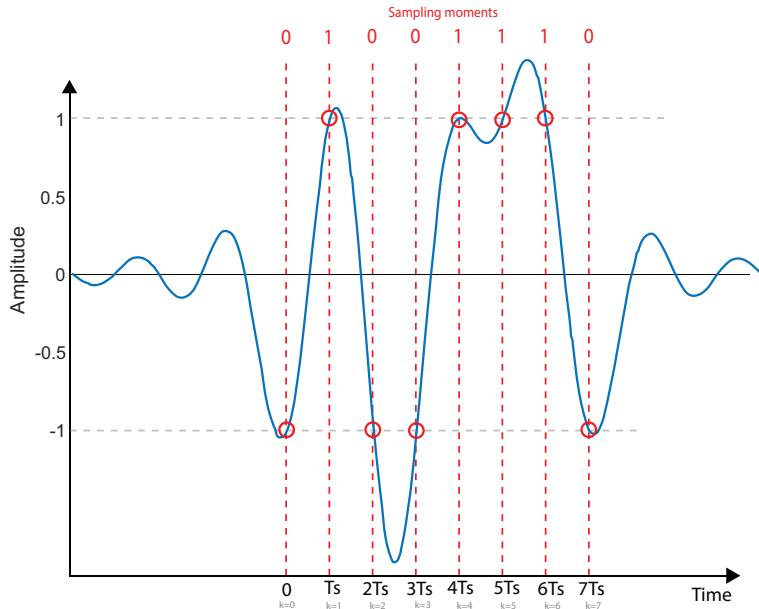


Figure 63: Actual final overlapped sinc pulses waveform representing letter N in binary, with sampling moments  $kT_s$ . The image shows that exactly on the  $kT_s$  sampling moments the value of the waveform is either 1 or -1, showcasing that sinc pulses are zero-ISI pulses

We can observe in the final waveform composed of the sinc pulses (Fig. 63), that at exactly the sampling moments  $kT_s$ , the value of the waveform is either 1 or -1. This shows that ideal sinc pulses are zero-ISI pulses because the sinc pulse transmitted on time 0 is only non-zero (contains an amplitude of -1) on time 0. In times  $T_s, 2T_s, 3T_s, \dots, 7T_s$ , it has zero amplitude, thus not interfering with the amplitude of the pulses on adjacent timeslots. And the same holds also for the sinc pulses transmitted at time  $T_s, 2T_s, \dots, 7T_s$ , they have zero amplitudes at other timeslots apart from their designated timeslots, and do not interfere with the amplitude of the pulses on the adjacent timeslots. This essentially makes sinc pulses zero-ISI.

Types of waveform pulses which satisfy zero-ISI criteria are summarized in Fig. 64.

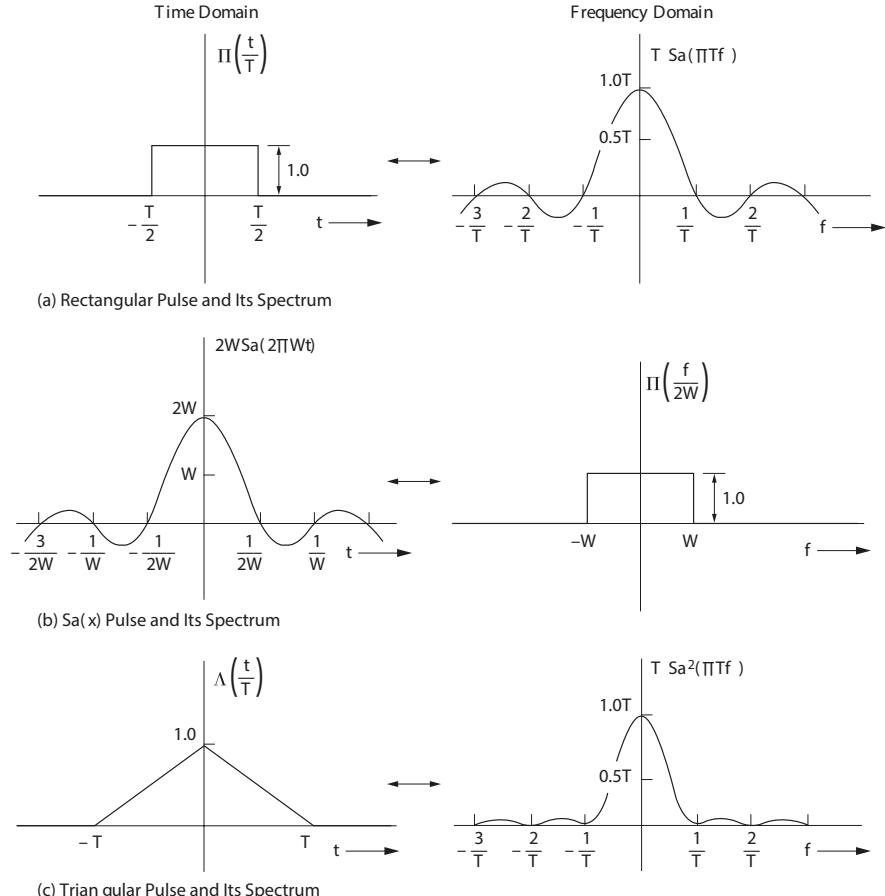


Figure 64: Ideal zero-ISI pulse waveforms. (a) Rectangular pulses, (b) sinc pulses, and (c) triangular pulses [2, ch. 2, p. 80, fig. 2-6]

Nevertheless, the issue with these pulses is that they are infeasible in practice. To get perfect no-ISI pulses, we need infinite bandwidth (for rectangular and triangular pulses) or we need negative and infinite time (for sinc pulses), both of which are not practical in real systems. Nevertheless, these pulses provide a basis for the general zero-ISI pulse waveform, which has a raised-cosine spectrum (RACOS), which is realizable and used in real systems, that minimizes the effects of ISI and has a relatively rectangular bandwidth.

## 8.6 Raised cosine-rolloff Nyquist filtering

Raised cosine-rolloff Nyquist filtering is used to generate pulses that minimize ISI. The transfer function of such a filter is

$$H_e(f) = \begin{cases} 1 & \text{for } |f| \leq f_1 \\ \frac{1}{2}[1 + \cos(\frac{\pi(|f|-f_1)}{2f_\Delta})] & \text{for } f_1 < |f| < B \\ 0 & \text{for } |f| \geq B \end{cases} \quad (108)$$

The transfer function spectrum is plotted in Fig. 65

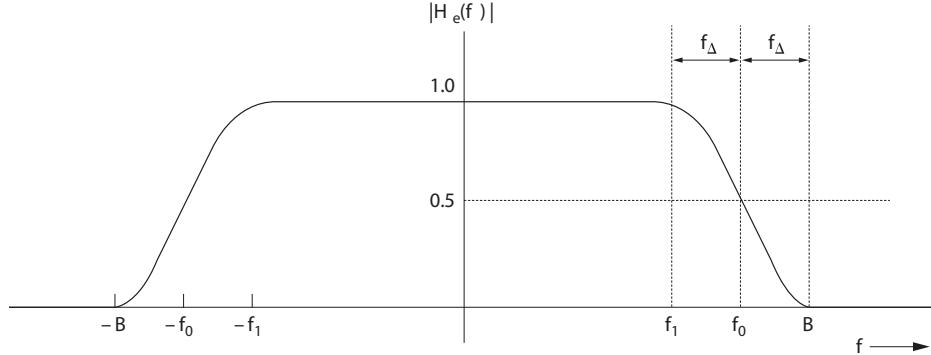


Figure 65: Raised cosine-rolloff Nyquist filter characteristics. [2, ch.3.6, p.211, fig. 3-25]

$B$  is the absolute bandwidth with the parameters

$$f_\Delta = B - f_0 \quad (109)$$

and

$$f_1 = f_0 - f_\Delta \quad (110)$$

where  $f_0$  is called the 6-dB bandwidth of the filter. Furthermore, we define the rolloff factor to be

$$r = \frac{f_\Delta}{f_0} \quad (111)$$

Which is the steepness of the filter. For  $r = 0$ , the filter is an ideal rectangular filter. Furthermore, bandwidth can also be defined as

$$B = f_0(1 + r) \quad (112)$$

The envelope of the filter decays with a factor of  $\frac{1}{t^3}$

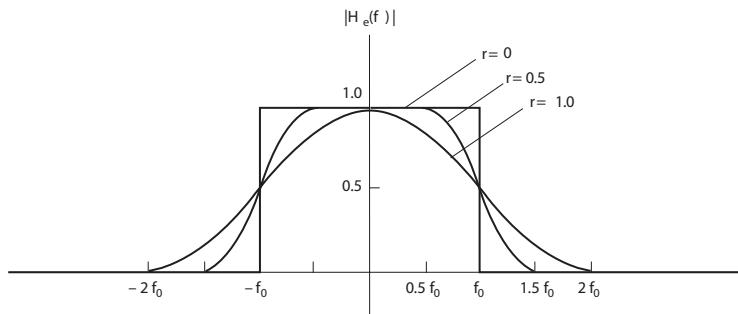


Figure 66: Raised cosine-rolloff Nyquist filter magnitude frequency response by varying  $r$  [2, ch.3, p.212 (3-26)]

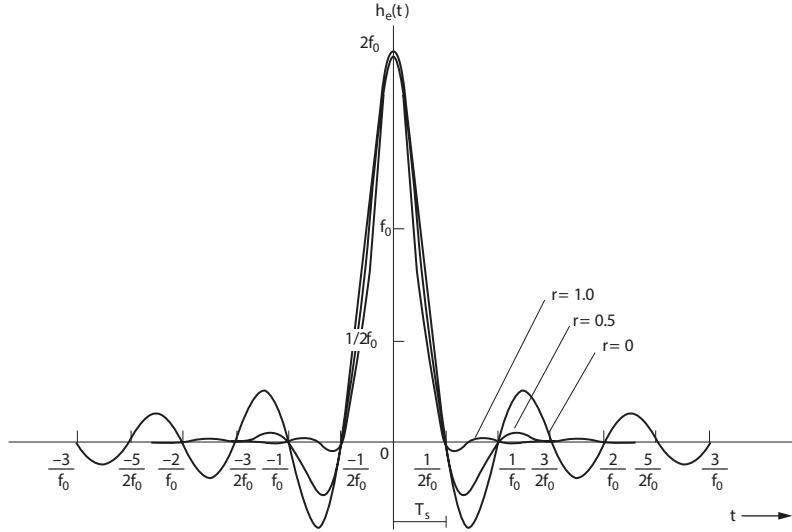


Figure 67: Impulse response  $h_e(t)$  of the raised cosine filter with varying  $r$  [2, ch.3, p.212 (3-26)]

Such a filter is obtained by convolving a band-limited cosine and a rectangular specturm filter (visualized in Fig. 68)

$$H_e(f) = \Pi\left(\frac{f}{2f_0}\right) * \left(\cos\left(\frac{\pi f}{2f_\Delta}\right) \cdot \Pi\left(\frac{f}{2f_\Delta}\right)\right) \quad (113)$$

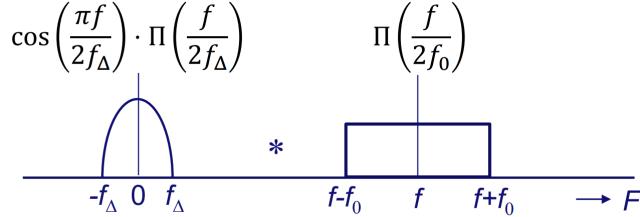


Figure 68: Convolution between a bandlimited cosine spectrum and a rectangular spectrum

By taking the inverse Fourier transform we can get the time-domain equation of raised-cosine pulses

$$h_e(t) = F^{-1}[H_e(f)] = 2f_0 \text{sinc}(2f_0 t) \left[ \frac{\cos(2\pi f_\Delta t)}{1 - (4f_\Delta t)^2} \right] \quad (114)$$

The symbol period of a raised cosine is  $\frac{1}{2}f_0$ . Where the maximum symbol rate for no ISI is

$$D_{max} = 2f_0 \quad (115)$$

To intuitively demonstrate the necessity of raised-cosine filtering, consider a bandlimited channel (e.g., a long CAT-1 twisted-pair wire) with a maximum frequency range up to  $f_{cut}$ . When sending a square pulse with a bandwidth exceeding the channel's maximum rated bandwidth, higher frequency components are attenuated, causing pulse spreading, interfering in adjacent time slots and oscillations (see Fig. 69a). To address this, on the otherhand sending a raised-cosine filtered pulse with a 0.75 roll-off factor, fitting the channel's bandwidth, results in significantly reduced pulse spreading (Fig. 69b).

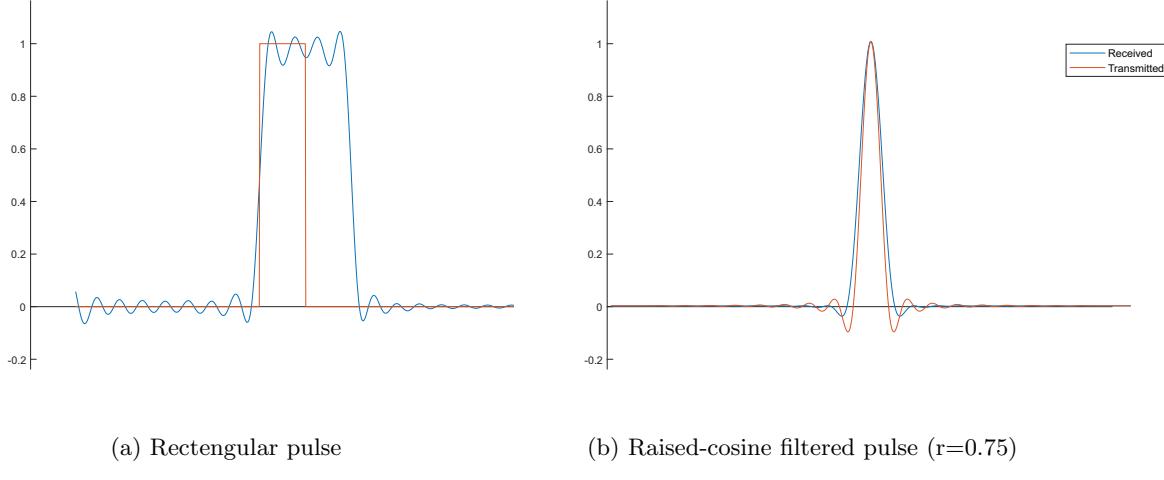


Figure 69: Figure demonstrating a rectangular pulse (a) and a raised-cosine filtered pulse (b) being sent through bandlimited channel. It can be noticed that a rectangular pulse spreads and interferes with adjacent timeslot (causing ISI), while a raised-cosine pulse almost doesn't spread at all, since it exactly fits the channels bandwidth. (*note that amplitude attenuation is neglected*)

Now to further observe the effects intuitively, suppose we want to transmit the letter 'N' through this channel (in binary 01001110) at a rate of 1 Mbps. We choose to transmit using unipolar-rz (URZ) with about 50% duty cycle line coding. The transmitted waveform can be observed in Fig. 70. Once this waveform passes through the channel, the received waveform at the receiving side is shown in Fig. 71. You may observe that the received waveform is completely distorted by ISI, in which the waveform does not look anymore like a unipolar-rz but rather like a unipolar-nrz, and if the receiver is tuned to decode unipolar-rz, the data may be completely corrupted.

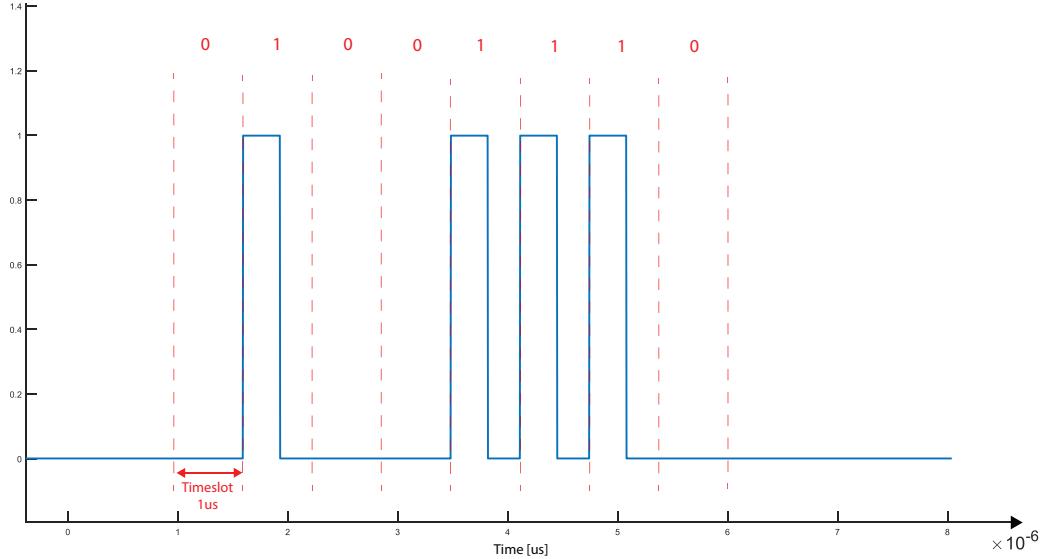


Figure 70: Transmitted pulse sequence using rectangular pulses, with unipolar-rz line coding

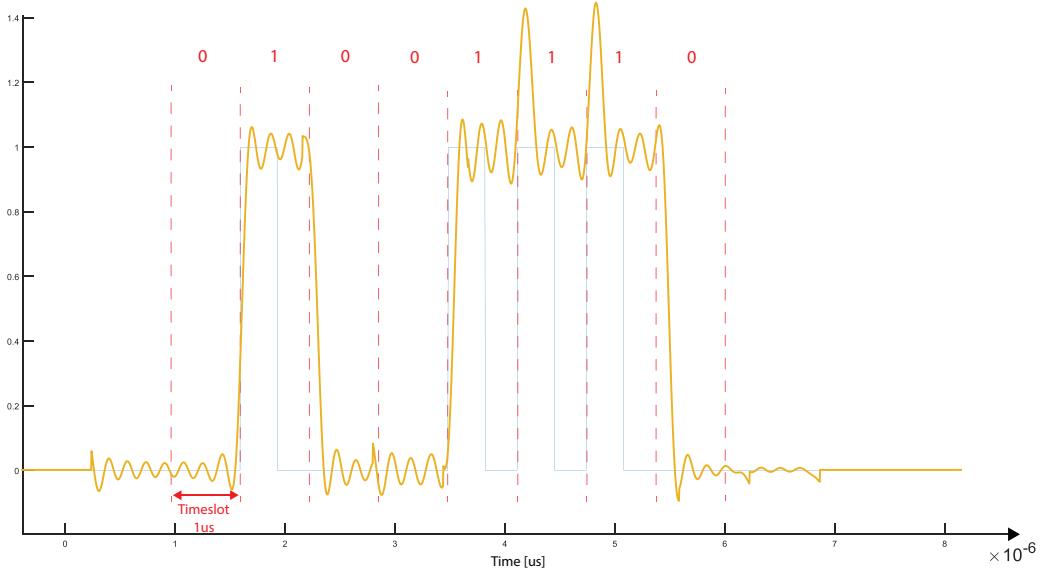


Figure 71: Received pulse sequence using rectangular pulses, with unipolar-rz line coding. It can be observed that due to ISI the pulses almost look like unipolar-nrz instead of unipolar-rz, which will cause corruption of data in the receiver side.

However, now we transmit the same data using the same line coding on the same channel, however, now we use raised-cosine pulses (with  $r=0.75$ ) instead of rectangular pulses. The received waveform is shown in Fig. 72. As can be observed, the raised cosine pulses much better represent the transmitted symbols in Fig. 70, and they have a value of 1 and 0 at the respective sampling times. The receiver, can then easily decode the data, at a rate of 1Mbps through this bandlimited channel.

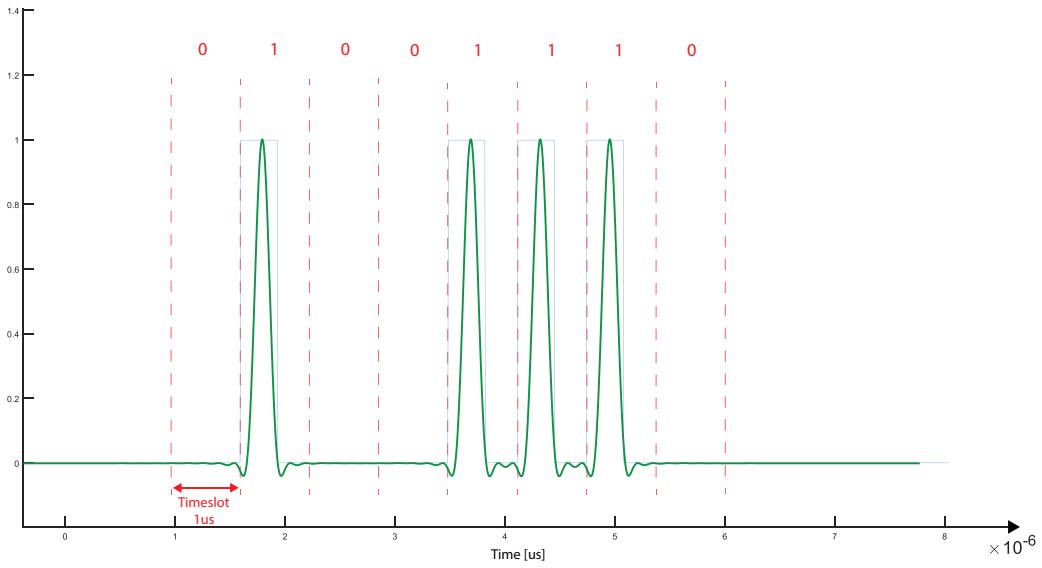
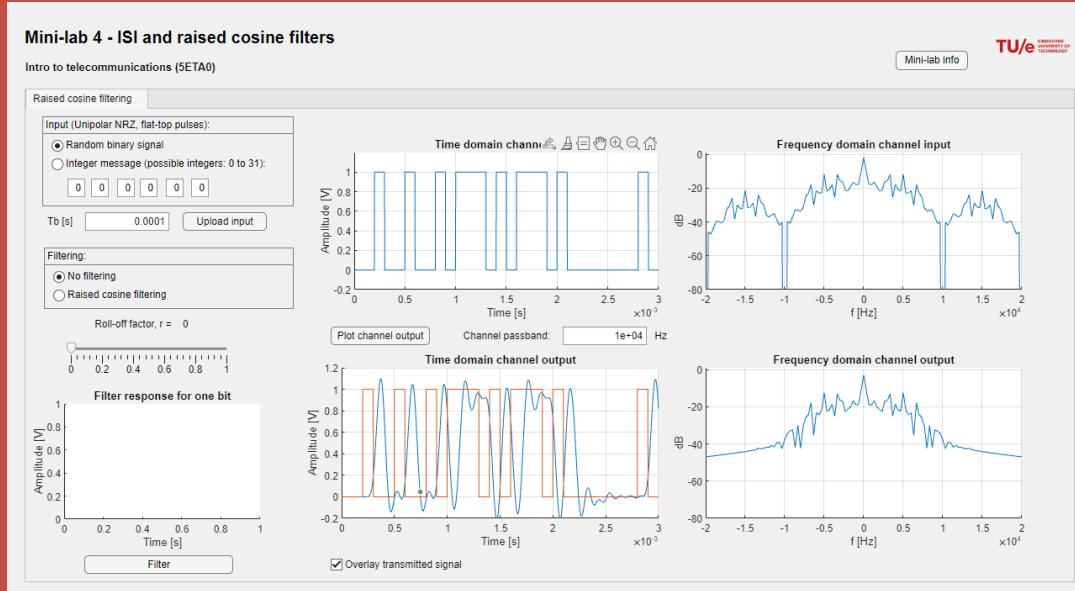


Figure 72: Received raised-cosine filtered pulse sequence, with unipolar-rz line coding. It can be observed that raised-cosine pulses do not get affected by the bandlimited channel, and they have a amplitude of 1, and 0 at their sampling instances, preserving the shape of unipolar-rz line coding

## Minilab exercise 8.2 - Raised-cosine filtering and ISI

This mini-lab exercise requires you to use Mini-lab 4 - ISI (Intersymbol interference)  
When you open Minilab 4 you will be confronted with the following view:



In this minilab exercise, you will experiment with the concept of how raised-cosine filtering prevents intersymbol interference.

- 1) Start by clicking '*Upload input*' to send the data input. Tick '*Overlay transmitted signal*' and then click '*Plot channel output*'
- 2) Now decrease the channel passband to 12000 Hz instead of 20000 Hz, and then click again on '*Plot channel output*' to see the effects of ISI. Can you observe how the pulses are smearing into adjacent timeslots?
- 3) Now to demonstrate the raised-cosine filtering, under the *filtering section*, click on raised cosine filtering, then click again on *Upload input*. Now you may tune your filter's time response by changing the roll-off factor. For every change, you should click '*filter*' and then '*plot channel output*' to observe the effects.
- 4) Choose a suitable roll-off factor, and apply it to the signal, plot it, and compare it to the unfiltered pulses, can you see how the raised cosine pulses minimise ISI?

## 8.7 Instruction exercises - ISI

The solutions to these exercises may be found under the page **5ETC0 Canvas Page Modules** → **Week 5** → **F. Intersymbol Interference**

**Exercise 12)** (*Video solution available*) A transmission system has an ideal (rectangular) transfer function with a bandwidth of 5 MHz.

- a) How many symbols per second can be transmitted over this channel and what is the pulse shape with which this can be accomplished?
- b) What is the maximum bit rate for transmission if Manchester NRZ binary signals are applied?
- c) What are the benefits and drawbacks of using Manchester NRZ line coding?

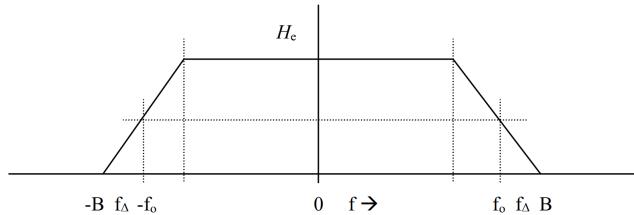
By means of a Nyquist filter, Manchester NRZ pulses are now converted into pulses having a "raised cosine" spectrum, with a roll-off factor  $r$ .

- d) How does the roll-off factor  $r$  influence the bandwidth of the signal and the detection in the receiver?

**Exercise 13)** (*Video solution available*) A baseband transmission system has an equivalent transfer function  $H_e(f)$  with a trapezoidal shape outlined:

$$H_e(f) = \begin{cases} 1 & \text{for } |f| < f_0 - f_\Delta \\ \frac{f_0 + f_\Delta - |f|}{2f_\Delta} & \text{for } f_0 - f_\Delta \leq |f| \leq B \\ 0 & \text{for } |f| > B \end{cases} \quad (116)$$

with  $B = f + f_\Delta$ .



- a) Explain why this is a "Zero-ISI" Nyquist filter

A PCM signal will be transferred with this baseband system. For the PCM signal applies:  
*sampling frequency*  $f_s = 4 \text{ kHz}$

*6 bits per sample*

For the transmission system applies:

$f_o = 12 \text{ kHz}$

$B = 16 \text{ kHz}$

- b) Determine the symbol rate and the number of levels per symbol that has to be applied.

The spectral power density of the noise at the input of the PCM receiver is  $No/2 = 10^{-7} \text{ W/Hz}$ , and the received signal power is 50 mW.

- c) Determine the signal to noise ratio ( $S/N$ )<sub>out</sub> at the output of the PCM receiver

- d) What is the improvement (in dB) of the signal-to-noise ratio of the PCM system with respect to analog transmission of the signal?

We reduce the number of bits per sample and adjust the bandwidth of the system accordingly.

- e) Calculate the resulting change of  $(S/N)_{out}$  and explain why this is an improvement or deterioration
- f) Calculate the pulse shape  $h_e(t)$ , which corresponds with the transfer function  $H_e(f)$ .

**Problem 14) (Video solution available)** A binary symbol is encoded as a waveform  $w(t)$  for which holds

$$s = 1 \Rightarrow w(t) = \Pi\left(\frac{2t}{T_b}\right) \quad (\text{rectangular pulse with duration } T_b / 2)$$

$$s = 0 \Rightarrow w(t) = -\Pi\left(\frac{2t}{T_b}\right)$$

- a) How would you call this form of signaling (polar, unipolar, bipolar or Manchester / NRZ or RZ?)

A random sequence of symbols  $s(k)$  is encoded as  $w(t - kTb)$  with  $Tb = 50\mu s$  and offered (unfiltered) to a transmission system.

- b) Give the expression for the power spectrum of the signal, make a sketch and indicate characteristic points.

The overall transfer of the transmission system is such that at the input of the detector, the pulse shape has a "raised-cosine" Nyquist spectrum with a roll-off factor of 0.2.

- c) Give the correct expression for the transfer function of the system, in other words, determine the parameters  $f_0$ ,  $f_1$ ,  $f_\Delta$ , and the total bandwidth B.

**Exercise 15) (Video solution available)** A binary symbol  $s$  is encoded as a waveform  $w(t)$  with:

$$w(t) = \begin{cases} \pm \cos(\pi t / \tau_s) & \text{for } -\tau_s/2 \leq t \leq \tau_s/2 \\ 0 & \text{elsewhere} \end{cases} \quad (117)$$

- a) What is the frequency range that covers the spectrum  $S(f)$  of this waveform? Explain your answer.

A series of these symbols with  $\tau_s = 150 \mu s$ , is sent unfiltered at a rate of 10 ksymbols/s. Detection takes place in the receiver at the center of the symbol interval by means of a 'sample and hold' circuit and a decision detector.

- b) Does inter-symbol interference (ISI) occurs when detecting? (hint: make a sketch)
- A "raised-cosine" Nyquistfilter is designed for the transmission of symbols which have a repetition frequency of 10 kHz. The roll-off factor is specified as  $r = 0.3$
- c) Give the correct expression for the transfer function of the filter, in other words, determine the parameters  $f_1$ ,  $f_\Delta$  and B.

**Exercise 16) (Video solution available)** With a "sample and hold" A/D converter an analog input signal is converted into a "flat-topped" PAM signal. The sampling frequency

is 25 kHz. The PAM signal is quantized into 4096 discrete levels. The levels are digitally encoded and the information is transmitted using multi-level polar RZ symbols. The number of levels per symbol is  $L = 8$ , and the "duty cycle" of the RZ pulses is 50

- a) What is the bit rate  $R$  of the information?
- b) What is the symbol rate (baud rate)  $D$  of the digital line signal?
- c) What is the null bandwidth of the digital (RZ) signal?

By means of a Nyquist filter, the RZ pulses are now converted into pulses having a "raised cosine" spectrum, with a roll-off factor  $r = 0.3$ .

- d) What is the required transmission bandwidth?

## 9 Information theory

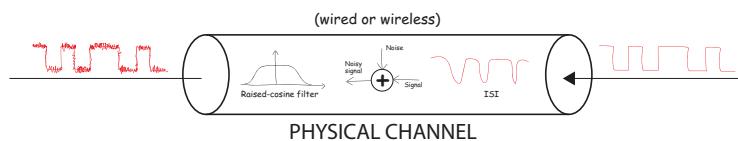
### 9.1 Learning objectives

Students completing this chapter should have learned:

1. Understand the fundamental limit for channel capacity based on the Shannon-Hartley Theorem.
2. Can calculate the maximum theoretical capacity of a channel based on the bandwidth and the SNR.
3. Understand the difference between power limited and bandwidth limited modes of operation for a communication channel.
4. Can calculate a parity bit for a simple case of digital word transmission.
5. Understands the concept of Hamming distance and can use Hamming(7,4) coding to detect and correct single bit errors.

### 9.2 Motivation

This section first introduces a maximum theoretical upper bound on the total amount of information (in bits) that can be sent reliably through a channel, given the bandwidth and SNR. This upper bound is known as *Shannon–Hartley capacity theorem*, and is a very important theorem in the field of digital communications. As we have observed, nearly every signal, when transmitted through a channel, undergoes corruption from noise, resulting in a probability of error (or bit flip) greater than zero. If the chance that the incoming digital data will contain a bit of flip due to noise is around 1 in 1 million (*So  $P_e = 10^{-6}$  which is typical in wireless systems*), and our data rate is 1 Mbps, this means that every second there will be a bit flip on the incoming data. So one of the questions that information theory deals with is: how can we detect in the receiver this bit flip and fix it? Thus, this section presents error detection and error correction schemes, which are used almost in every digital communication or even digital storage system.



### Shannons capacity theorem

Figure 73: Shannon's channel capacity theorem - Topic map location

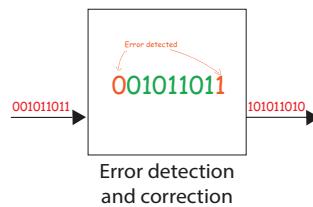


Figure 74: Error correction and detection schemes - Topic map location

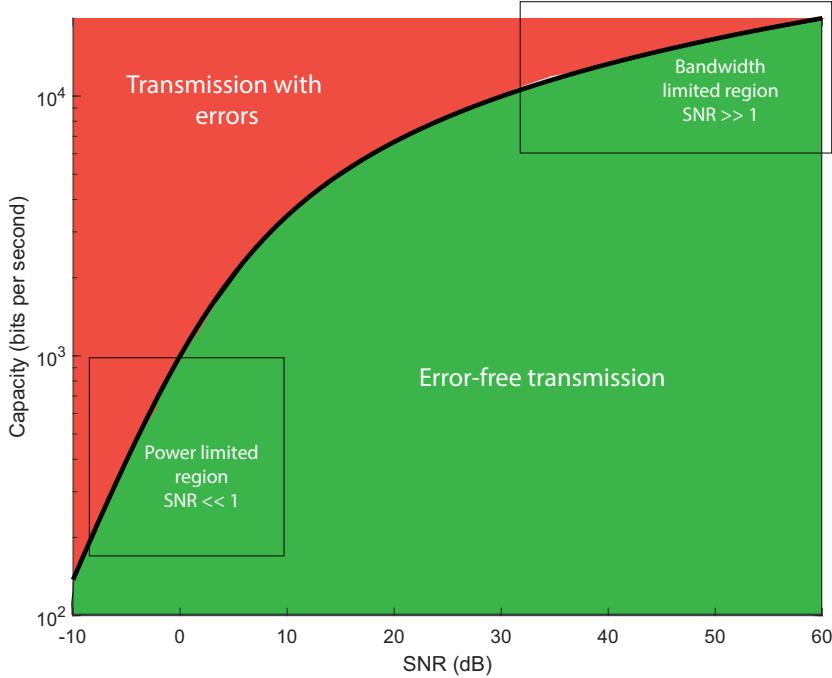


Figure 75: Shannon-Hartley channel capacity equation plotted for SNR -10 dB to 60 dB, with a channel bandwidth of 1 kHz

### 9.3 Shannon–Hartley channel capacity theorem

The Shannon-Hartley capacity theorem defines the upper limit for the rate at which information in bits per second can be sent, *error-free*, over a communication channel with a certain bandwidth and signal power in the presence of noise. The equation is defined as

$$C = B \log_2(1 + SNR) \quad [\text{bits/s}] \quad (118)$$

where  $C$  is the channel capacity (bits/s),  $B$  is the channel bandwidth (Hz) and  $SNR$  is the signal-to-noise ratio of the channel. The maximum spectral efficiency is then defined as

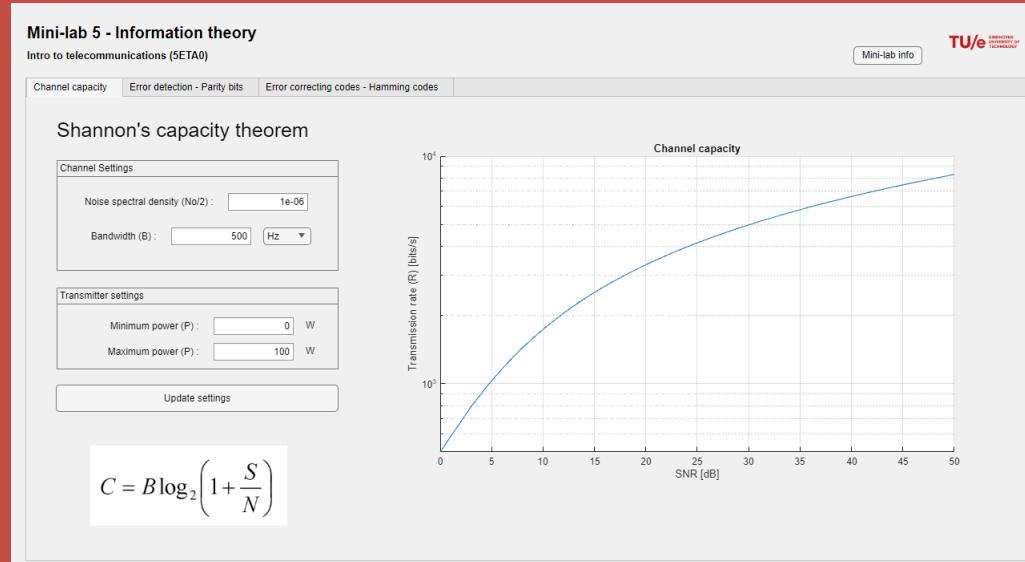
$$\eta_{max} = \frac{C}{B} = \log_2(1 + SNR) \quad (119)$$

Equation 118 is plotted in Fig. 75.

We observe a linear relationship between bandwidth and data rate, as well as a logarithmic relationship between signal power and baud rate. In Fig. 75, two distinct regions are observed: the **bandwidth-limited region** and the **power-limited region**. The power-limited region occurs when the signal-to-noise ratio (SNR) is significantly less than 1 ( $SNR \ll 1$ ). Within this domain, the channel capacity is constrained by insufficient SNR. Consequently, the data rate can experience exponential growth by increasing the SNR. On the other hand, the bandwidth-limited region is applicable when SNR is notably greater than 1 ( $SNR \gg 1$ ). In this scenario, there is an abundance of SNR, but the limiting factor becomes the available bandwidth in the channel. Irrespective of how much SNR is available, the data rate reaches saturation, and increasing the SNR does not increase the capacity by much. We conclude that in this region more bandwidth is essential to facilitate the transmission of additional information through the channel.

## Minilab exercise 9.1 - Shannons channel capacity theorem

This mini-lab exercise requires you to use Mini-lab 5 - Information Theory on MATLAB.



In this minilab exercise, we will explore the Shannon-Hartley channel capacity theorem and its implications.

- 1) Leave the default settings. Try to identify where is the SNR-limited region in the channel capacity plot, and the bandwidth-limited region.
- 2) Change the noise spectral density (No/2, under the channel settings tab), to a value of 0.1, and click update settings. What do you see now? Is the system operating in a bandwidth-limited region or SNR-limited region?
- 3) Assume you are designing a telecommunication system for a smartphone, and you want to achieve a bitrate of 10 Mbps ( $10^7$  bits per second). The channel noise spectral density between the smartphone and the base station is on average  $10^{-8}$ . As a smartphone designer, the transmit power is constrained up to a maximum of 0.25 watts. What bandwidth do you need to achieve this transmission rate considering the transmission power limitations?
- 4) What happens with the SNR range? Does it decrease or increase when you increase the bandwidth? Explain why this happens.

## 9.4 Error detection and correction schemes

Error detection and correction are crucial in digital communications, especially when using channels like wireless ones where errors (such as bitflips) due to noise are common (*why would they be more common in wireless communication compared to wired communication?*).

Being able to detect if and when bits have been received with an error will allow us to discard wrong data and not use it for further processing. In this information theory section, we will explore a simple error detection method called parity bit checking. Having a good error detection scheme, will allow us to find faulty messages and retry to send the data across the channel and hope that no errors occur this time. This can be very expensive and does not allow for high data transmission rates which we need for most applications. It can also severely congest the network and cause many other undesirable effects (*The exact impact of packet re-transmission is not included in this course*).

A better way would be if we could somehow create self-correcting data, which gets self-corrected at the receiver, such that we do not have to resend it again. Using the principles of parity bit checking, we will develop an error correction scheme known as *Hamming codes*.

### 9.4.1 Parity bit checking - (Error detection)

A parity bit is a binary digit that is added to a binary message to ensure that the total number of ones in the data is even or odd. The purpose of the parity bit is to provide a simple error-checking mechanism, primarily in communication systems and data storage. Suppose we want to send the following message:

$$01001110 \quad (120)$$

which represents the letter 'N' in binary. The sender agrees with the receiver on a parity format, which is either *odd parity* or *even parity*. Suppose they agree on *odd parity*. The sender now counts the number of 1s in the message (4 in the case of letter N), and attaches a new bit equal to 1 to the message, called the parity bit, which makes the total count of 1s to 5, which is *odd parity*. The message now becomes

$$010011101 \quad (121)$$

Note that if the sender and receiver were to agree on *even parity*, then the attached parity bit would be 0 instead of 1, since already the number of ones is 4, which is an even number.

Now the sender sends the message above with odd parity, and the message is received as

$$000011101 \quad (122)$$

On the receiver side, the second bit has flipped. Now the receiver counts the number of 1s of this received message, and the number is 4, which is even, however, both the sender and receiver have agreed that the number of ones must add up to odd number (*odd parity*). In this way, the receiver now knows that a bit has flipped on the way, and hence has successfully *detected an error on the received data*. This is how the simple parity bit checking works for error detection.

*Question: What would be the actual limitation of such a system?. Try to answer before moving to the next page*

#### 9.4.2 Limitations of parity bit error detection

Parity bit checking is a very simple error detection methodology. However, it is also limited, suppose two bits flip instead of one

$$0100\textcolor{red}{0}111 \quad (123)$$

The number of ones in this message is now 5, which is an odd number. The receiver would compute this, and would see that it is a odd parity, and would assume that the message has no errors, however, that message represents the letter 'G' instead of the intended letter 'N', but the error would go undetected. Simple parity bit checking is used in simple wired communication systems such as USB between two devices. More advanced error detection schemes are CRC (cyclic redundancy checks), computing checksums, etc. However these are outside the scope of this course.

## Minilab exercise 9.2 - Parity bit checking

This mini-lab exercise requires you to use Mini-lab 5 - Information Theory on MATLAB.

The screenshot shows the 'Error detection - Parity bits' tab selected in the navigation bar. The interface is divided into three main sections: Transmitter, Channel, and Receiver. In the Transmitter section, the parity bit is set to 'Even', the bit stream is '100101', and the transmitted bitstream is '1001011'. In the Channel section, the number of bits to flip is set to '1'. In the Receiver section, the parity bit is also set to 'Even', the received bitstream is '0001011', and the button 'Analyze for errors' is highlighted in red with the message 'ERROR DETECTED!' displayed.

In this exercise, you will experiment with simple parity bit error detection. In the channel settings, you may choose the number of bits you want to flip so that the received message contains errors. Furthermore, on the transmitter, you may change the message (in bits, called the bit stream) and the parity bit type (*even or odd*). On the receiver settings, you may also change the parity bit type, and everytime you want the system to analyze for errors please click on the '*analyze for errors*' button. Furthermore, for any change made on the transmitter and channel settings, please always click the *send* button to resend the data to the receiver.

- 1) Leave default settings, but change the parity bit type on receiver side to odd parity, and click to analyze errors. Is the error analysis result correct? If not, explain why
- 2) Keep the settings the same as in the previous question, but now increase the nr of bits flipped to 1, resend the message, and perform error checking. Is the error analysis result correct? If not, explain why
- 3) Now change the parity bit settings of the receiver to *even* type. Resend the message, and perform error-checking. Is the error analysis result now correct?
- 4) Increase now the number of bits flipped to 2. Resend the message, and perform error-checking. Is the error analysis result correct? If not, explain what is happening in an intuitive way

### 9.4.3 Hamming codes - (Error correction and detection)

Hamming codes are a method for both detecting and correcting errors in a transmitted message. Suppose we want to send a binary message of length  $m$ . To ensure error detection and correction, we add  $k$  parity bits, where the total number of bits in the transmitted sequence becomes  $n = m + k$ . These parity bits are positioned strategically to allow the detection and precise correction of errors in the message.

For example, let's consider we want to transmit the message **1010**. We will add three parity bits to this message, resulting in a total of seven bits. This setup is known as the Hamming(7,4) code, where  $n = 7$ ,  $m = 4$ , and  $k = 3$ .

In Hamming codes, the parity bits are placed at positions that are powers of 2: 1, 2, 4, 8, and so on. In our 7-bit Hamming coded message, the bits are arranged as follows:

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
$p_1$	$p_2$	$x_3$	$p_3$	$x_2$	$x_1$	$x_0$

Here,  $p_n$  represents the  $n$ -th parity bit, and  $x_n$  represents the  $n$ -th data bit. If we substitute the data bits with our message (1010), we obtain:

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
$p_1$	$p_2$	1	$p_3$	0	1	0

Next, we compute the parity bits one by one.

**1. First Parity Bit ( $p_1$ ):** The first parity bit  $p_1$  checks the bits at positions 1, 3, 5, and 7. It ensures that the number of ones in these positions is even.

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
$p_1$	$p_2$	1	$p_3$	0	1	0

Here, the sequence of bits at positions 1, 3, 5, and 7 is {1, 1, 0, 0}, with a total of one 1, which is odd. To make the count even, we set  $p_1 = 1$ . The updated sequence is:

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
1	$p_2$	1	$p_3$	0	1	0

**2. Second Parity Bit ( $p_2$ ):** The second parity bit  $p_2$  checks the bits at positions 2, 3, 6, and 7. It ensures that the number of ones in these positions is even.

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
1	$p_2$	1	$p_3$	0	1	0

Here, the sequence of bits at positions 2, 3, 6, and 7 is {0, 1, 1, 0}, with two ones, which is even. Hence, we set  $p_2 = 0$ . The updated sequence is:

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
1	0	1	$p_3$	0	1	0

**3. Third Parity Bit ( $p_3$ ):** The third parity bit  $p_3$  checks the bits at positions 4, 5, 6, and 7. It ensures that the number of ones in these positions is even.

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
1	0	1	$p_3$	0	1	0

Here, the sequence of bits at positions 4, 5, 6, and 7 is {0, 0, 1, 0}, with one 1, which is odd. To make the count even, we set  $p_3 = 1$ . The updated sequence is:

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
1	0	1	1	0	1	0

Thus, the final Hamming-coded message is:

$$1011010 \quad (124)$$

## Error Detection and Correction

Once the receiver receives a message, they compute the parity bits for the received sequence, and compare them with the received parity bits. If any discrepancy is found, an error is detected.

Suppose one bit of the transmitted message flips, turning 1011010 into:

$$101101\textcolor{red}{1} \quad (125)$$

The receiver takes the message from the bits in positions 3, 5, 6, and 7, which are:

$$1011 \quad (126)$$

It computes the parity bits for this sequence, resulting in:

$$0110011 \quad (127)$$

Next, it performs an XOR operation between the received sequence and the computed sequence:

$$1011011 \oplus 0110011 = \textcolor{green}{1101000} \quad (128)$$

The result of the XOR operation reveals the position of the flipped bit, as the bits at the parity bit locations that are set to 1 correspond to the binary representation of the error position. In this case, the error occurs at bit position 7, which is where the flip happened.

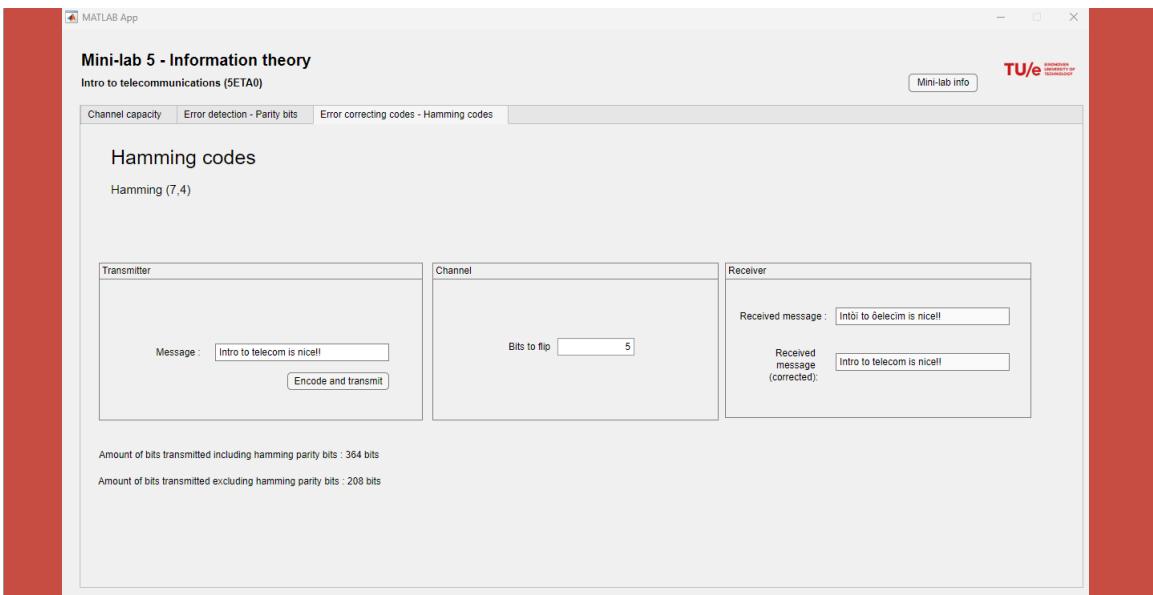
If no error occurs, the received and computed sequences will match exactly, and their XOR will result in a sequence of all 0s, indicating no error.

### Pitfalls of hamming codes

While hamming codes excel at correcting a single flipped bit within a sequence block, they fall short when faced with multiple flipped bits. Despite this limitation, hamming codes can still detect the presence of multiple errors, surpassing the limitations of simple parity bit checking. It's worth noting that, hamming codes were the pioneers in error correction codes. Modern advanced telecommunications systems like 5G rely on more sophisticated schemes such as LDPC (low-density parity check) codes and polar codes, which are beyond the scope of this course.

### Minilab exercise 9.3 - Hamming codes (error correction)

This mini-lab exercise requires you to use Mini-lab 5 - Information Theory on MATLAB.



In this exercise, we will explore hamming(7,4) codes. On the transmitter settings, you may write the message you want to transmit. On the channel settings, you may choose the number of bits to flip, and on the receiver side, you will see how the message was received and how it transformed after hamming error correction was performed. Moreover, on the bottom left, you may see the total number of bits of the transmitted sequence and only the number of bits of the message.

- 1) Change the number of bits flipped to 1, and re-transmit the message, to observe the effects of error correction.
- 2) Increase the number of bits flipped to 3, and re-transmit the message. Is the message correctly shown to the receiver? If so, can you explain how is this possible when in theory, hamming code sequences can correct only one bit flip?
- 3) Now keep increasing the number of bit's flipped and re-transmit the message. Is it possible to correctly decode the message after a lot of bit flips have occurred?

## 9.5 Instruction exercises - Information Theory

The solutions to these exercises may be found under the page **5ETC0 Canvas Page Modules** → **Week 5** → **G. Information theory**

**Exercise 17)** (*Video solution available*) A music signal  $w(t)$  is band-limited to  $B = 20$  kHz. The signal is sampled with a frequency  $f_s = 50$  kHz. The samples are uniformly quantized and coded into a PCM signal of 12 bits/sample, and sent as unipolar binary NRZ symbols over a channel with an ideal low-pass characteristic and is disturbed by additive white Gaussian noise. In the receiver detection of the digital signal and restoration of the music signal takes place. The required minimum signal to noise ratio of the recovered signal is 60 dB.

- a) What is the maximum allowable bit error rate when the required signal-to-noise ratio is achieved?
- b) What is the minimum signal to noise ratio of the received binary signal?
- c) If the bandwidth of the channel is limited, how still distortion-free transmission of the PCM signal can be achieved?

**Exercise 18)** (*Video solution available*) A data signal is transported by means of a transmission system. The system operator has purchased from the government 1GHz of bandwidth and wants to maximize the capacity he can deliver to his customers.

- a) Assuming the absolute theoretical limit of capacity can be reached and that the only noise at the receiver is thermal noise. What is the required signal to noise at the receiver to allow a bit rate of 4Gb/sec?
- b) What is the signal power ( $P$ ) needed? (Given that noise power is  $4.14 \cdot 10^{-12}$  Watt)
- c) Would doubling the bandwidth double the bit rate? Motivate your answer
- d) How much more bandwidth is needed in order to increase the capacity to 5Gb/sec?

When implementing the system the engineers have chosen to use Bipolar RZ with a pulse width 50 % of the period.

- e) What would be the actual bandwidth used by the system if we send data at a rate of 1Gb/sec?
- f) Would the bandwidth needed change if you reduce to pulse width to 25 %? By how much?

# 10 Amplitude, Frequency and Phase Modulation

## 10.1 Learning objectives

Students completing this chapter should have learned:

1. Understand the difference between baseband and passband modulation.
2. Understand the concept of up and down conversion using a mixer.
3. Can illustrate a simple circuit for up/down conversion to/from passband to base band and calculate the right frequency for the local oscillator.
4. Can draw the time evolution of amplitude modulated signals with and without a carrier (AM and DSB-SC AM).
5. Can draw the Frequency spectrum of an AM modulated signal (both AM and DSB-SC AM).
6. Can calculate the modulation percentage and efficiency as well as the PEP for AM modulated signals.
7. Can sketch and explain the operation principle of envelope and product detectors for AM modulated signals and their limitations.
8. Can sketch the time evolution and spectrum for FM/PM modulated signals.
9. Can calculate modulation index  $\beta_f$  for FM modulated signal.
10. Can apply Carson's rule to deduce bandwidth needs for an FM modulated signal.

## 10.2 Motivation

So far we have discussed the communication of information assuming it is all done at the same frequency range in which it is sampled and quantized. For practical reasons, the use of baseband signaling is not desirable. While this can still be considered for point to point connections using a dedicated channel (think of a wire connecting two circuits) when transmitting over free-space, if all users are transmitting their data in baseband they will all occupy the same part of the spectrum, and communication will not be possible due to interference. We chose therefore to give different users different frequencies to reduce undesirable cross-talk. When superimposing the data on a designated carrier (passband modulation) we are then also given a choice of how to convert the information for passband transmission. In this chapter we discuss in detail the use of amplitude and phase or frequency modulation.

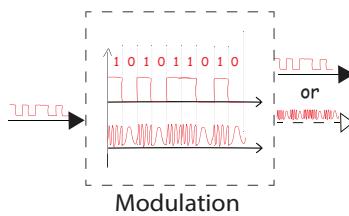


Figure 76: Topic map location - AM/FM/PM

## 10.3 Baseband vs Bandpass

First of all, let us realize what the difference is between a baseband and a bandpass signal. Baseband refers to the spectrum of the transmitted signal being centred at 0Hz. This is practical for cases where we are not interested in how we affect the overall frequency spectrum of the environment, such as is the case for optical fiber communication. Since the

transmission is shielded from the outside world, it does not matter in fact how much of the available spectrum we use. In the case of open-air transmission, this is very different, since there are strict regulations on what part of the frequency spectrum one can transmit on (See table 1, for common frequency band regulations) .

Source	Allocated Frequency band (MHz)
Longwave BCB (EU)	0.150–0.285
AM BCB (EU)	0.153–0.279
AM BCB (US)	0.530–1.710
Amateur	1.8–1.9
Citizens band	26–28
Amateur	28–30
Land mobile	30–50
Amateur	50–54
TV low VHF	54–88
Land mobile (EU)	66–88
FM BCB (EU)	87.5–108
FM BCB (US)	88–108
Aircraft	108–136
Land mobile (J)	142–174

Table 1: Common frequency allocations

To solve this, we upconvert the signal, thus re-centering from 0Hz to a carrier frequency. This means starting from a spectrum such as depicted here (Baseband):

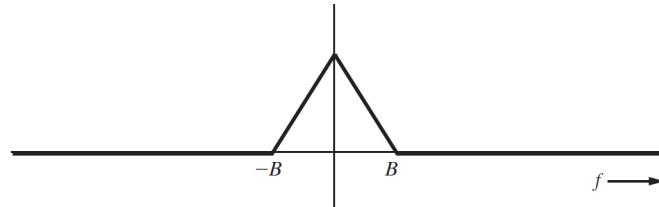


Figure 77: Spectrum of a baseband signal (centered around 0 Hz)

into one that is re-centered, with representations for both positive and negative frequency components (Bandpass).

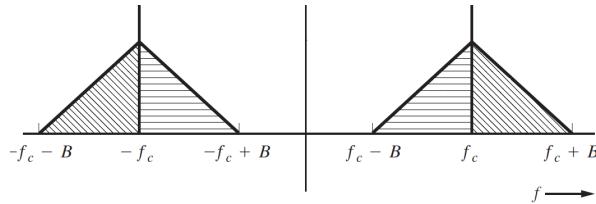


Figure 78: Spectrum of a bandpass signal (centered around a carrier frequency  $f_c$ )

### 10.3.1 Up and down conversion

To shift a signal between baseband and bandpass, a circuit comprising a mixer, a local oscillator (LO), and a filter is employed. The mixer multiplies the input signal by the LO signal. This operation is based on the trigonometric identity:

$$\cos(\alpha) \cos(\beta) = \frac{1}{2} \cos(\alpha + \beta) + \frac{1}{2} \cos(\alpha - \beta). \quad (129)$$

This relation shows that the multiplication results in two frequency components: one at the sum  $(\alpha + \beta)$  and one at the difference  $(\alpha - \beta)$  of the original frequencies. A filter is then used to select the desired component, thus achieving either upconversion (shifting to a higher frequency band) or downconversion (shifting to baseband). The typical topology to achieve up and down conversion is shown in Fig. 79.

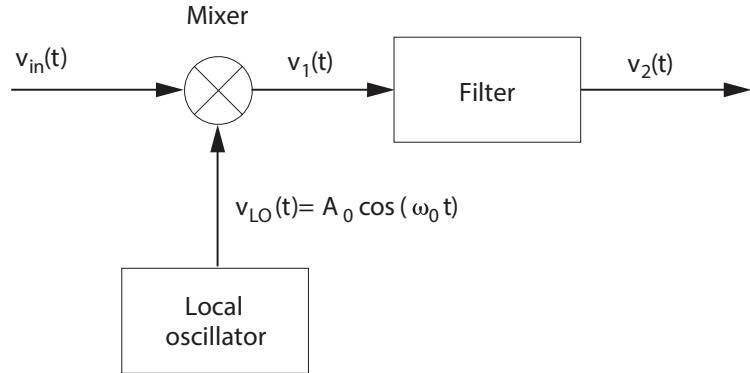


Figure 79: Up and down conversion circuit topology [2, ch.4-11, p.291]

The local oscillator (LO) generates a stable, sinusoidal waveform at a fixed frequency, which is the carrier frequency  $f_c$ .

## 10.4 Amplitude Modulation (AM)

Amplitude Modulation (AM) is one of the simplest modulation schemes used in analog communications, but can also be used to modulate digital signals. In AM, the amplitude of a high-frequency carrier signal is varied in proportion to a lower-frequency baseband message signal  $m(t)$ . Essentially, the message (e.g. music) waveform is encoded on the high-frequency carrier's signal amplitude

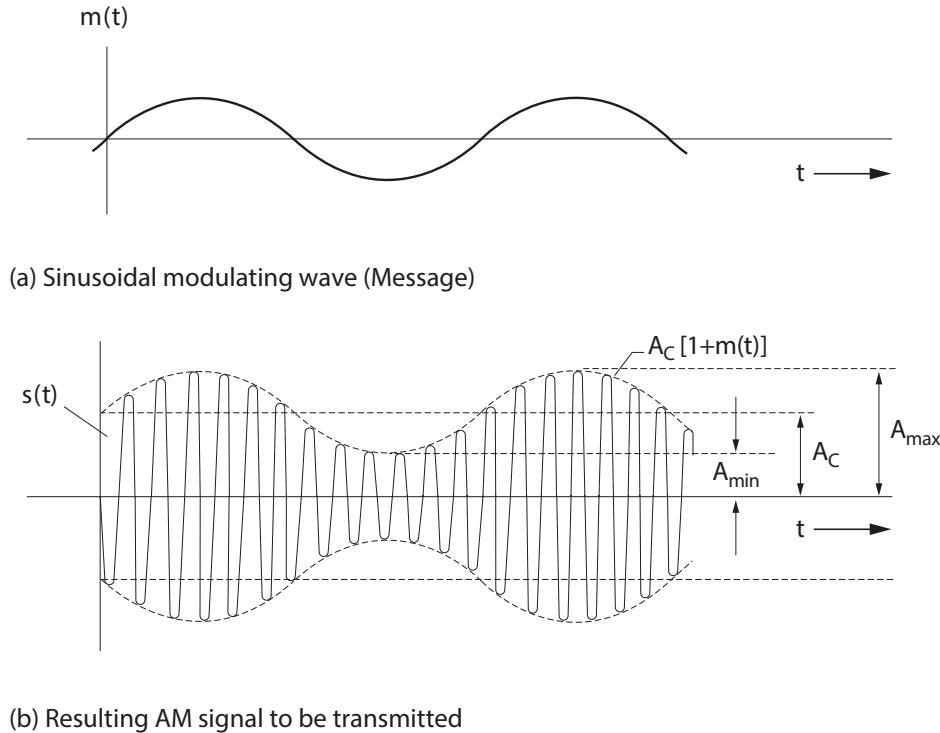


Figure 80: Typical AM Waveform, where  $m(t)$  is a low-frequency sine wave.

With the notations:

- $g(t)$ , the envelope of the signal (dashed line on Fig. 80b)
- $A_c$ , the carrier amplitude
- $m(t)$ , the modulating signal (essentially the message we want to transmit)
- $s(t)$ , the AM signal

The equations for the transmitted signal are defined as:

$$g(t) = A_c[1 + m(t)] \quad (130)$$

$$s(t) = g(t) \cos(\omega_c t) \quad (131)$$

You may observe how exactly the modulating signal affects the amplitude of the waveform at any given moment. It then follows that the maximum value that the modulated signal will reach, is equal to  $A_c (1 + \max[m(t)])$ , and the minimal value  $A_c (1 + \min[m(t)])$ . Hence we define the two as being  $A_{\max}$  and  $A_{\min}$ .

#### 10.4.1 Percentage of modulation

The percentage of modulation (or modulation index expressed as a percentage) quantifies the degree to which the amplitude of a carrier is varied by the baseband message signal  $m(t)$ . Essentially, it provides a measure of how much the information signal influences the overall amplitude of the transmitted wave. We define the percentage of modulation as

$$\%mod = \frac{\max(m(t)) - \min(m(t))}{2} \cdot 100\%. \quad (132)$$

Where the percentage of **positive modulation** is given by  $\max[m(t)] \cdot 100\%$ , and that of **negative modulation** by  $-\min[m(t)] \cdot 100\%$ .

#### 10.4.2 Modulation efficiency

Modulation efficiency in AM scheme measures how effectively the transmitted power is used to convey information. In an unsuppressed AM system, the total transmitted power is divided between the carrier and the sidebands, but only the sidebands actually contain the message information. The modulation efficiency  $\eta_{mod}$  is defined as

$$\eta_{mod} = \frac{\langle m^2(t) \rangle}{1 + \langle m^2(t) \rangle} \cdot 100\% \quad (133)$$

where  $\langle m^2(t) \rangle$  is the mean-squared value of the modulating signal  $m(t)$ . For example, for a sinusoidal signal, the mean squared-value  $\langle m^2(t) \rangle$  can be computed as

$$\langle m^2(t) \rangle = \frac{1}{T} \int_0^T \sin^2(\omega t) dt = \frac{1}{T} \int_0^T \left( \frac{1}{2} - \frac{1}{2} \cos(2\omega t) \right) dt = \frac{1}{2} \quad (134)$$

#### 10.4.3 Peak envelope power

The peak envelope power represents the maximum instantaneous power that the modulated signal's envelope reaches over a period of time. This value is crucial because it determines the highest power level that a transmitter must handle. The peak envelope power  $P_{PEP,norm}$  is defined as

$$P_{PEP,norm} = \frac{1}{2R} [\max(|g(t)|)]^2 = \frac{A_c^2}{2R} [1 + \max[m(t)]]^2. \quad (135)$$

Where  $R$  is a load resistance. In power calculations for modulation systems, we assume a load  $R$  to serve as a reference for how the signal delivers power. In real circuits, the transmitter, amplifier, or antenna is connected to a load that has a certain impedance (often  $50 \Omega$  in radio systems).

#### 10.4.4 Spectrum of AM signals

Recalling the general equation for AM signals

$$s(t) = A_c [1 + m(t)] \cos(2\pi f_c t), \quad (136)$$

We can take its Fourier transform to observe how the spectrum of such signals will look like. By expanding eq. 136, we have

$$s(t) = A_c \cos(2\pi f_c t) + A_c m(t) \cos(2\pi f_c t) \quad (137)$$

We note that the first term is the carrier component and the second term is the sideband (where the message is contained).

- 1. **Carrier Component:** The carrier signal  $A_c \cos(2\pi f_c t)$  has the Fourier transform:

$$\mathcal{F}\{A_c \cos(2\pi f_c t)\} = \frac{A_c}{2} \delta(f - f_c) + \frac{A_c}{2} \delta(f + f_c). \quad (138)$$

- 2. **Sidebands:** The term  $A_c m(t) \cos(2\pi f_c t)$  modulates the carrier, resulting in shifted spectra (*Using the property that a multiplication in time domain is a convolution in frequency domain*):

$$\mathcal{F}\{A_c m(t) \cos(2\pi f_c t)\} = \frac{A_c}{2} M(f - f_c) + \frac{A_c}{2} M(f + f_c), \quad (139)$$

where  $M(f)$  is the Fourier transform of  $m(t)$ .

The overall spectrum  $S(f)$  is the sum of the carrier and sideband components:

$$S(f) = \frac{A_c}{2} \delta(f - f_c) + \frac{A_c}{2} \delta(f + f_c) + \frac{A_c}{2} M(f - f_c) + \frac{A_c}{2} M(f + f_c). \quad (140)$$

If the baseband message  $m(t)$  has a bandwidth  $B$  (i.e.,  $M(f) = 0$  for  $|f| > B$ ), then:

- The upper sideband extends from  $f_c - B$  to  $f_c + B$ ,
- The lower sideband extends from  $-f_c - B$  to  $-f_c + B$ .

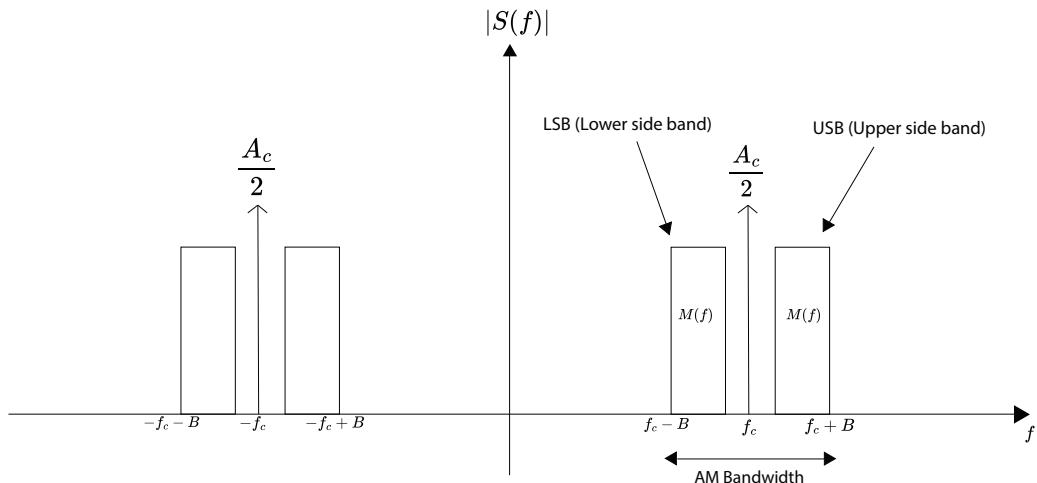


Figure 81: Spectrum of AM Signal assuming  $m(t)$  has a square bandwidth.

## 10.5 Double-Sideband Suppressed Carrier (DSB-SC) Modulation

Unlike conventional AM, where the carrier is transmitted along with the sidebands, DSB-SC is a variant of AM that eliminates the carrier, transmitting only the sideband components that actually carry the information, but at a bandpass centered around  $f_c$ . This absence of the carrier is clearly visible in the accompanying figure, where the signal's envelope fluctuates from its maximum value down to zero, demonstrating the direct influence of the message signal on the amplitude.

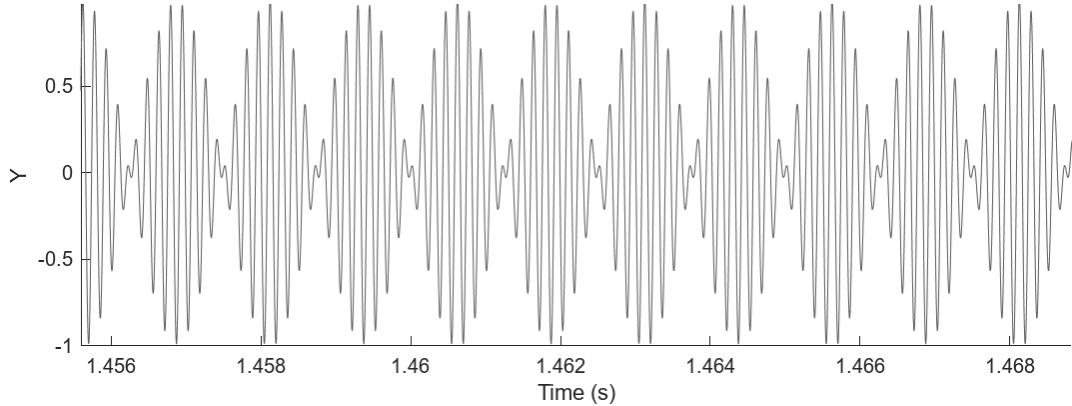


Figure 82: DSB-SC signal in time domain modulating a sine wave.

To express this mathematically, we begin by defining a scaled version of the message signal as

$$g(t) = A_c m(t) \quad (141)$$

where  $A_c$  is the carrier amplitude and  $m(t)$  represents the modulating signal. The transmitted DSB-SC signal is then given by

$$s(t) = A_c m(t) \cos(\omega_c t) = \frac{A_c}{2} m(t) (e^{j\omega_c t} + e^{-j\omega_c t}), \quad (142)$$

which illustrates that the modulation process results in two sidebands symmetrically positioned about the carrier frequency  $f_c$ .

Since the carrier is not present in the transmitted signal, the conventional concept of a **modulation percentage** becomes undefined (or conceptually infinite) because there is no constant carrier component to compare against. Nonetheless, this approach yields a **modulation efficiency** of 100%, as all the transmitted power is devoted solely to conveying the information contained in  $m(t)$ .

The efficiency of DSB-SC makes it particularly attractive in scenarios where power conservation is critical. However, this benefit is balanced by the need for coherent detection at the receiver. Specifically, the receiver must generate a carrier that is exactly synchronized in both frequency and phase with the transmitter's original carrier in order to recover the message signal effectively.

### 10.5.1 Spectrum of DSB-SC signals

Similarly to the spectrum of an unsuppressed AM signal, the difference is that the carrier component has an amplitude of 0. Hence, equation 140 becomes

$$S(f) = \frac{A_c}{2} M(f - f_c) + \frac{A_c}{2} M(f + f_c). \quad (143)$$

for DSB-SC modulation.

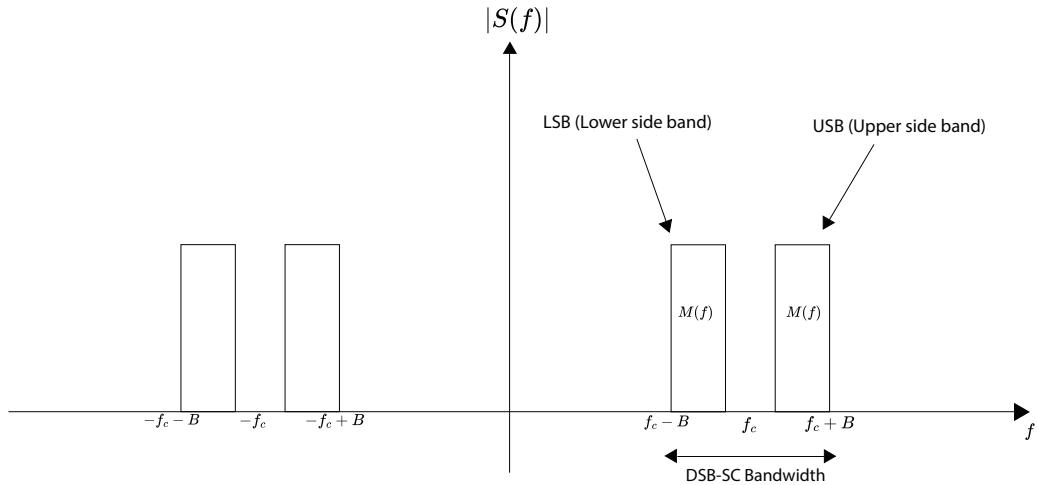


Figure 83: Spectrum of DSB-SC Signal assuming  $m(t)$  has a square bandwidth.

## 10.6 AM Detector circuits

In any communication system, once the AM modulated signal reaches the receiver, the first task is to extract (demodulate) the original message from the high-frequency carrier. This extraction is achieved by a detection circuit, and in this section we discuss two common methods for demodulation: the **envelope detector** and the **product detector**. Both approaches have distinct advantages and limitations, making them suitable for different signal conditions.

### 10.6.1 Envelope Detector

The envelope detector is one of the simplest circuits used in analog communication. Its operation is based on the observation that the amplitude variations (or “envelope”) of a modulated signal represent the original message signal. The circuit generally consists of:

- **A diode:** This component rectifies the incoming signal by allowing only one polarity (typically the positive half-cycle) to pass through.
- **An RC (resistor-capacitor) filter:** Following the diode, the capacitor charges and then discharges slowly through the resistor. This action smooths out the rapid fluctuations of the carrier and follows the slowly varying envelope of the signal.

Figure 84 illustrates a typical envelope detector circuit. When a modulated signal is applied, the circuit outputs a voltage that mirrors the envelope of the waveform, thereby recovering the baseband message. Figure 85 shows a typical result of this detection process.

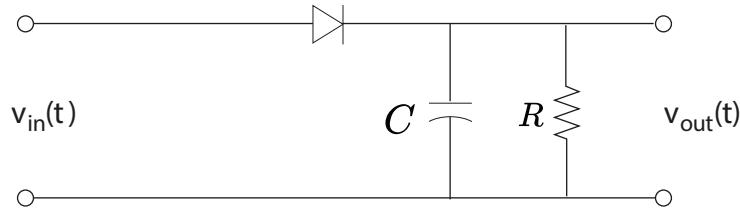


Figure 84: Envelope detector circuit: A simple detection circuit using a diode and an RC filter.

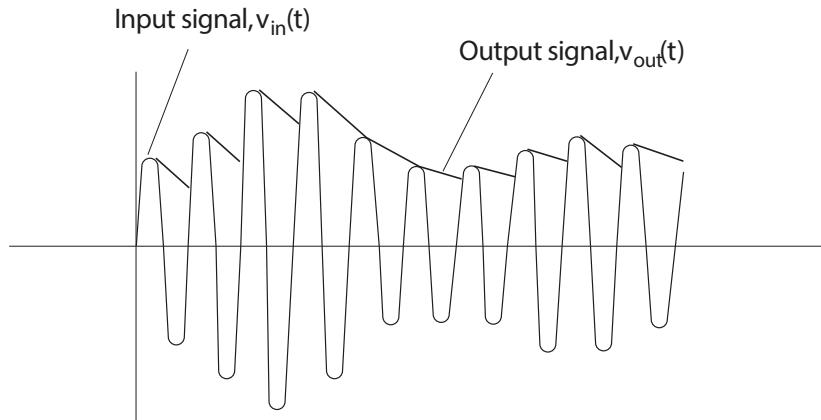


Figure 85: Output voltage from the envelope detector, showing the recovered message signal.

This method is very effective when the modulation index is within acceptable limits (typically less than 100%). However, if the signal is overmodulated (i.e., the negative modulation percentage exceeds 100%) or if the carrier is suppressed, the envelope detector may produce distorted outputs or fail entirely.

To illustrate this limitation, suppose we transmit a sine wave message using overmodulated AM (i.e., the negative modulation percentage exceeds 100%). Figure 86 shows the overmodulated AM signal.

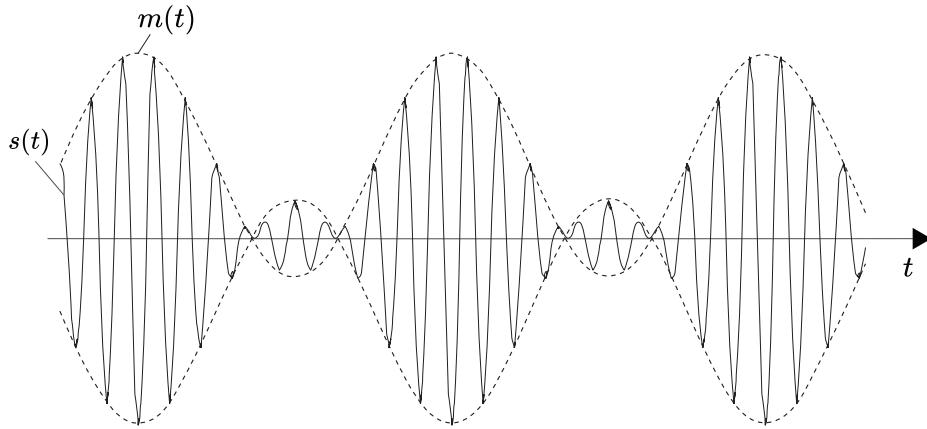


Figure 86: Illustration of overmodulated AM signal, carrying a sine-wave as a message  $m(t) = 1.5 \sin(2\pi ft)$

Upon receiving the overmodulated AM signal, we use an envelope detector to recover the original message signal, and the result of the original message can be seen in Fig. 87.

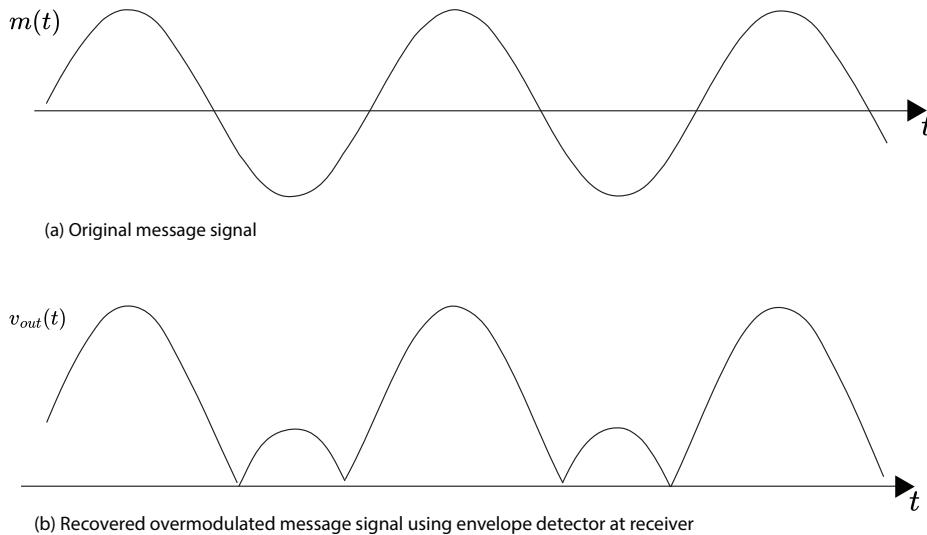


Figure 87: Figure illustrating the original message signal (a), and (b) the recovered message signal using envelope detector after being transmitted with an overmodulated AM. It can be seen that the recovered message does not anymore represent the original transmitted message; this shows the limitation of the envelope detector.

### 10.6.2 Product Detector

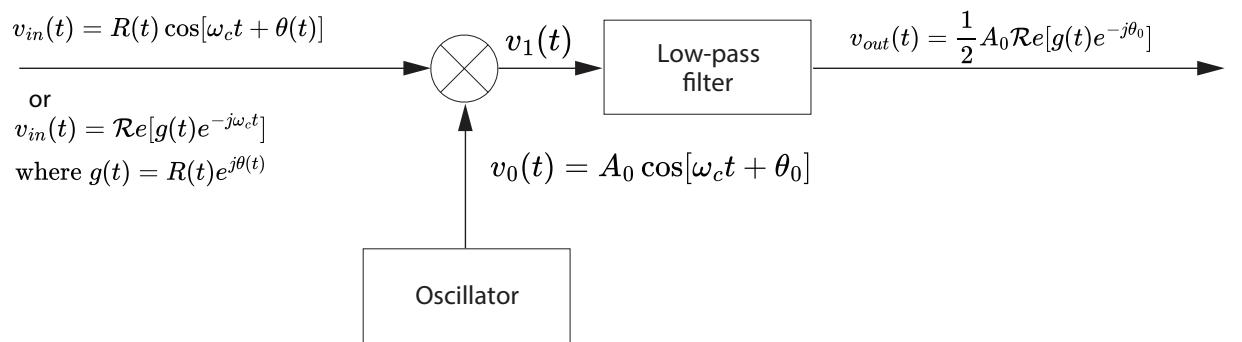
To overcome the limitations of the envelope detector, especially under conditions of over-modulation or carrier suppression, the **product detector** is used. Unlike the envelope de-

tector, the product detector actively multiplies the received modulated signal with a locally generated carrier that is synchronized in frequency and phase with the original transmitter's carrier. This process, known as coherent detection, "downconverts" the modulated signal back to baseband.

The product detector works as follows:

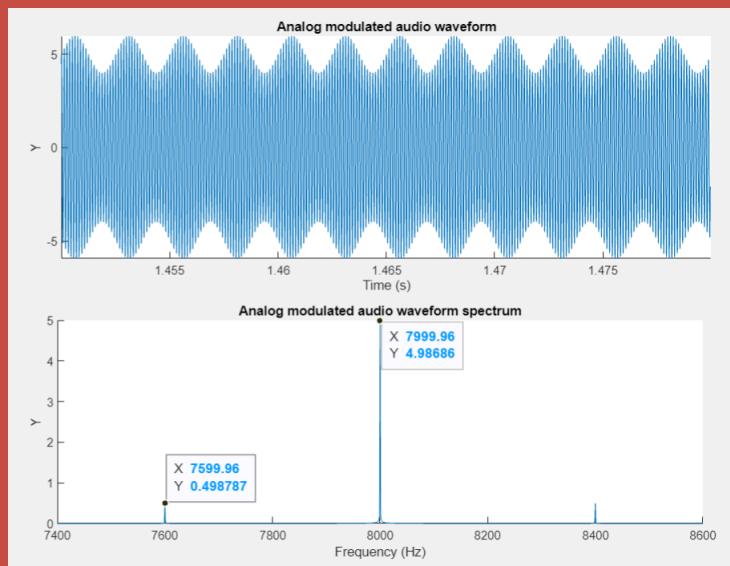
- **Mixing:** The incoming modulated signal is multiplied by a locally generated carrier signal. This multiplication produces two components: one at the sum of the frequencies and one at the difference.
- **Low-Pass Filtering:** A low-pass filter then removes the high-frequency (sum) component, leaving only the baseband (difference) component which contains the original message.

Figure 88 shows a typical product detector circuit. The key advantage of this method is its ability to recover the message accurately even when the modulation depth is extreme or when the carrier is suppressed.



## Minilab exercise 11.1

This mini-lab exercise requires you to use Mini-lab 6 - AM/FM on MATLAB.



Please firstly try to figure out the modulation index from the details given in the figure. Furthermore, use that information to obtain the modulation efficiency as well. What are your findings?

To verify this, play around with both the "Message" and "AM Transmitter" options within the Mini-lab. Provided you match the right input information, you should receive the correct values which you can then transform into modulation efficiency.

- 1) Think about what type of detector one could use to decode the information.
- 2) Now set the carrier amplitude to 0V. What do you observe in the transmitted signal? Can you still use the same method for detecting your transmission?

## Minilab exercise 11.2

This mini-lab exercise requires you to use Mini-lab 6 - AM/FM on MATLAB.

Now we will analyze the effect of noise on the system. Firstly, select the frequency of your message ton in the 'Message' tab. Furthermore, in the 'AM transmitter' tab select an  $f_c$  of 8kHz and  $A_c$  of 4V. Finally, go to the 'Channel' tab and enter a value of 0.0005 for  $N_0$  and click on 'Add noise AM'. In the 'AM Receiver' tab, use the 'Envelope detector' functions to decode the message.

- 1) Can you successfully hear the input tone?
- 2) What happens if you repeat the steps, increasing  $N_0$ ?
- 3) Use the parameters from Minilab exercise 11.1, and try to complete all the steps. What do you observe?

## 10.7 Frequency and phase modulation

While amplitude modulation (AM) is a robust method for transmitting information, it encodes the message solely in the amplitude of the carrier wave. For improved fidelity

and enhanced resistance to noise, modern communication systems often employ frequency modulation (FM) or phase modulation (PM). In both FM and PM the amplitude of the transmitted signal remains constant, while the information is carried in the variations of frequency or phase, respectively.

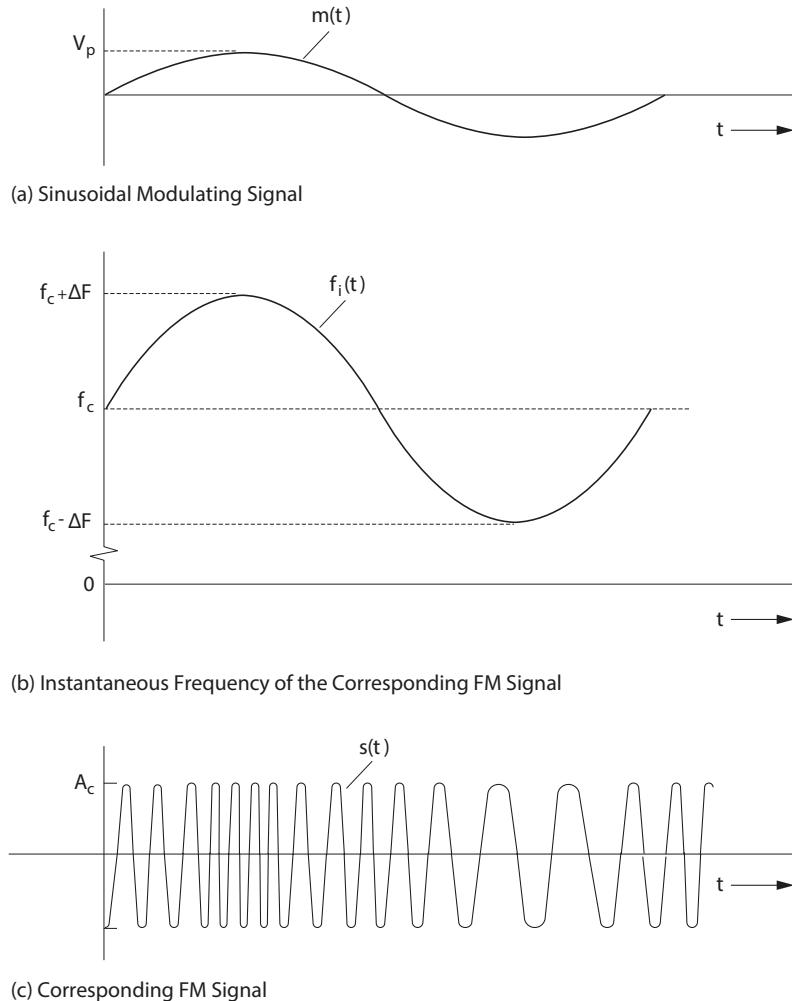


Figure 89: FM with a sinusoidal baseband modulating signal [2]

In **frequency modulation (FM)**, the instantaneous frequency of the carrier is varied in proportion to the baseband message signal. In other words, the carrier frequency deviates from its nominal value depending on the amplitude of the modulating signal. This technique provides improved noise immunity and a higher quality of signal reproduction. The complex envelope of an FM signal is given by:

$$g(t) = A_c e^{j\theta(t)}, \quad (144)$$

so that the real, transmitted bandpass signal can be written as:

$$s(t) = \Re\{g(t)e^{j\omega_c t}\} = A_c \cos(\omega_c t + \theta(t)), \quad (145)$$

where  $A_c$  is the constant carrier amplitude and  $\theta(t)$  is a time-varying phase term that encodes the information.

In contrast, **phase modulation (PM)** directly varies the phase of the carrier in proportion to the modulating signal. Although both FM and PM maintain a constant amplitude, they differ in how the modulating signal  $m(t)$  influences the phase term  $\theta(t)$ . In FM the phase is determined by the time integral of the message signal, while in PM the phase is directly proportional to the modulating signal.

For FM the phase is given by:

$$\theta(t) = D_f \int_{-\infty}^t m(\tau) d\tau, \quad (146)$$

and the instantaneous frequency deviation can be expressed as:

$$f_d(t) = \frac{1}{2\pi} \frac{d\theta(t)}{dt} = \frac{D_f}{2\pi} m(t), \quad (147)$$

where  $D_f$  is the frequency deviation constant. For PM the relationship is even simpler:

$$\theta(t) = D_p m(t), \quad (148)$$

with  $D_p$  representing the phase sensitivity. In both cases the carrier amplitude  $A_c$  is maintained constant, which minimizes the effect of amplitude noise.

### 10.7.1 Frequency and phase modulation index

We begin by defining the peak frequency deviation, which represents the maximum change in the instantaneous frequency of the carrier signal due to modulation. In frequency modulation (FM), the instantaneous frequency is determined by the time derivative of the phase,  $\theta(t)$ , divided by  $2\pi$ . Thus, the peak frequency deviation is defined as

$$\Delta F = \max \left\{ \frac{1}{2\pi} \frac{d\theta(t)}{dt} \right\} = \frac{1}{2\pi} D_f \max [m(t)],$$

where  $D_f$  is the frequency deviation constant and  $m(t)$  is the modulating signal. This equation shows that the maximum shift in frequency is directly proportional to the highest amplitude of the modulating signal.

The frequency modulation index is an important parameter that compares this deviation to the bandwidth of the modulating signal. It is defined as

$$\beta_f = \frac{\Delta F}{B} \quad (149)$$

where  $B$  represents the bandwidth of the modulating signal. This index,  $\beta_f$ , indicates how significant the frequency variations are relative to the frequency content of the message.

Conversely, in phase modulation (PM), the information is conveyed through direct variations in the phase of the carrier. The peak phase deviation is defined as the maximum value of the phase shift from the unmodulated carrier, given by

$$\Delta\theta = \max\{\theta(t)\} = D_p \max [m(t)], \quad (150)$$

where  $D_p$  is the phase sensitivity constant. For phase modulation, the modulation index is simply the peak phase deviation:

$$\beta_p = \Delta\theta \quad (151)$$

### 10.7.2 Carson's Rule

In frequency and phase modulation, the signal spectrum becomes quite complex due to the nonlinear relationship between the modulated signal  $g(t)$  and the message  $m(t)$ . This nonlinearity prevents us from deriving a simple, exact formula linking the spectrum  $G(f)$  to  $M(f)$ .

To address this, we use a rule-of-thumb known as Carson's rule, which estimates the bandwidth that contains most of the modulated signal's power. Specifically, Carson's rule states that approximately 98% of the total power is confined within the bandwidth

$$B_T = 2(\beta + 1) \cdot B = 2\Delta F + 2B \quad (152)$$

Here,  $\beta$  is the modulation index (which can be either the frequency or phase modulation index),  $\Delta F$  is the peak frequency deviation, and  $B$  is the bandwidth of the modulating signal. This rule provides a practical means of estimating the spectral occupancy of a modulated signal.

## 10.8 Frequency Spectrum of FM and PM

To analyze the frequency content of FM and PM signals, we begin with the complex envelope of the modulated signal. This envelope is given by

$$g(t) = A_c e^{j\theta(t)} = A_c e^{j\beta \sin(\omega_m t)}, \quad (153)$$

where  $A_c$  is the carrier amplitude,  $\beta$  is the modulation index, and  $\omega_m$  is the angular frequency of the modulating signal.

The transmitted signal is then obtained by combining this envelope with the carrier frequency:

$$s(t) = \Re\{g(t)e^{j\omega_c t}\} = \Re\{A_c e^{j(\beta \sin(\omega_m t) + \omega_c t)}\} = A_c \cos[\omega_c t + \beta \sin(\omega_m t)], \quad (154)$$

where  $\omega_c$  is the carrier angular frequency.

In the frequency domain, the envelope  $g(t)$  has a Fourier transform represented by a sum of delta functions weighted by Bessel functions:

$$G(f) = A_c \sum_{n=-\infty}^{\infty} J_n(\beta) \delta(f - n f_m), \quad (155)$$

and consequently, the spectrum of the modulated signal is

$$S(f) = \frac{1}{2} [G(f - f_c) + G(-f - f_c)], \quad (156)$$

with  $f_c = \omega_c/(2\pi)$  being the carrier frequency and  $f_m = \omega_m/(2\pi)$  the modulating frequency. Here,  $J_n(\beta)$  denotes the Bessel function of the first kind of order  $n$ .

Intuitively, the Bessel functions  $J_n(\beta)$  describe how the modulation index  $\beta$  determines the distribution of power among the spectral components (sidebands) of the modulated signal. Each  $J_n(\beta)$  quantifies the amplitude of the  $n$ th sideband, located at frequencies offset from the carrier by  $n f_m$ . When  $\beta$  is small, most of the power is concentrated in the carrier ( $n = 0$ ) and in the first-order sidebands ( $n = \pm 1$ ); as  $\beta$  increases, the power spreads over a larger number of sidebands.

The Bessel functions is defined by the integral

$$J_n(\beta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{j(\beta \sin \theta - n\theta)} d\theta, \quad (157)$$

and they satisfy the symmetry relation

$$J_{-n}(\beta) = (-1)^n J_n(\beta). \quad (158)$$

Figures below illustrate a visualization of several first-kind Bessel functions and the corresponding values of  $J_n(\beta)$  as a function of  $\beta$ :

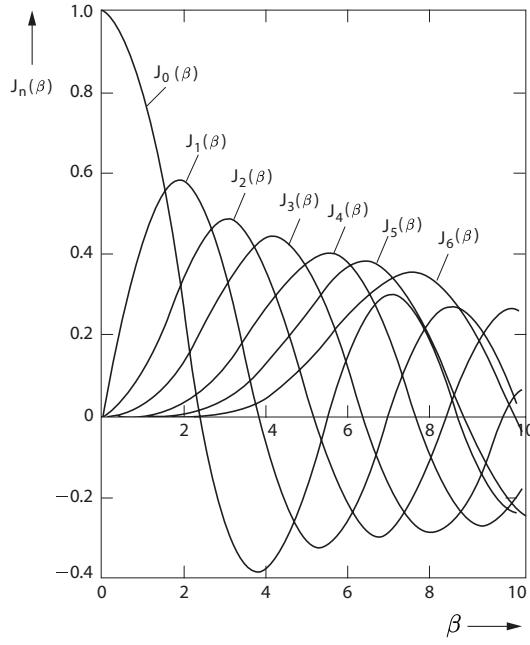


Figure 90: Visualization of various first-kind Bessel functions. [2]

		FOUR-PLACE VALUES OF THE BESSSEL FUNCTIONS $J_n(b)$											
$n \backslash b:$	0.5	1	2	3	4	5	6	7	8	9	10		
0	0.9385	0.7652	0.2239	- 0.2601	- 0.3971	- 0.1776	0.1506	0.3001	0.1717	- 0.09033	- 0.2459		
1	<u>0.2423</u>	0.4401	0.5767	0.3391	- 0.06604	- 0.3276	- 0.2767	- 0.004683	0.2346	0.2453	0.04347		
2	0.03060	<u>0.1149</u>	0.3528	0.4861	0.3641	0.04657	- 0.2429	- 0.3014	- 0.1130	0.1448	0.2546		
3	0.002564	0.01956	<u>0.1289</u>	0.3091	0.4302	0.3648	0.1148	- 0.1676	- 0.2911	- 0.1809	0.05838		
4		0.002477	0.03400	<u>0.1320</u>	0.2811	0.3912	0.3576	0.1578	- 0.1054	- 0.2655	- 0.2196		
5			0.007040	0.04303	<u>0.1321</u>	0.2611	0.3621	0.3479	0.1858	- 0.05504	- 0.2341		
6				0.001202	0.01139	0.04909	<u>0.1310</u>	0.2458	0.3392	0.3376	0.2043	- 0.01446	
7					0.002547	0.01518	0.05338	<u>0.1296</u>	0.2336	0.3206	0.3275	0.2167	
8						0.004029	0.01841	0.05653	<u>0.1280</u>	0.2235	0.3051	0.3179	
9							0.005520	0.02117	0.05892	<u>0.1263</u>	0.2149	0.2919	
10								0.001468	0.006964	0.02354	0.06077	<u>0.1247</u>	0.2075
11									0.002048	0.008335	0.02560	0.06222	<u>0.1231</u>
12										0.002656	0.009624	0.02739	0.06337
13										0.003275	0.01083	0.02897	
14										0.001019	0.003895	0.01196	
15											0.001286	0.004508	
16												0.001567	

Figure 91: Values of first-kind Bessel functions as a function of  $\beta$ . [2]

By combining these results, we obtain a complete frequency domain representation of FM and PM signals. This detailed spectral structure, governed by the Bessel functions, not only shows how the modulation index  $\beta$  influences the sideband amplitudes but also ties directly into Carson's rule. Carson's rule uses this insight to estimate the effective bandwidth  $B_T$  required for the modulated signal, considering that significant spectral components are spread over a range determined by  $\beta$ .

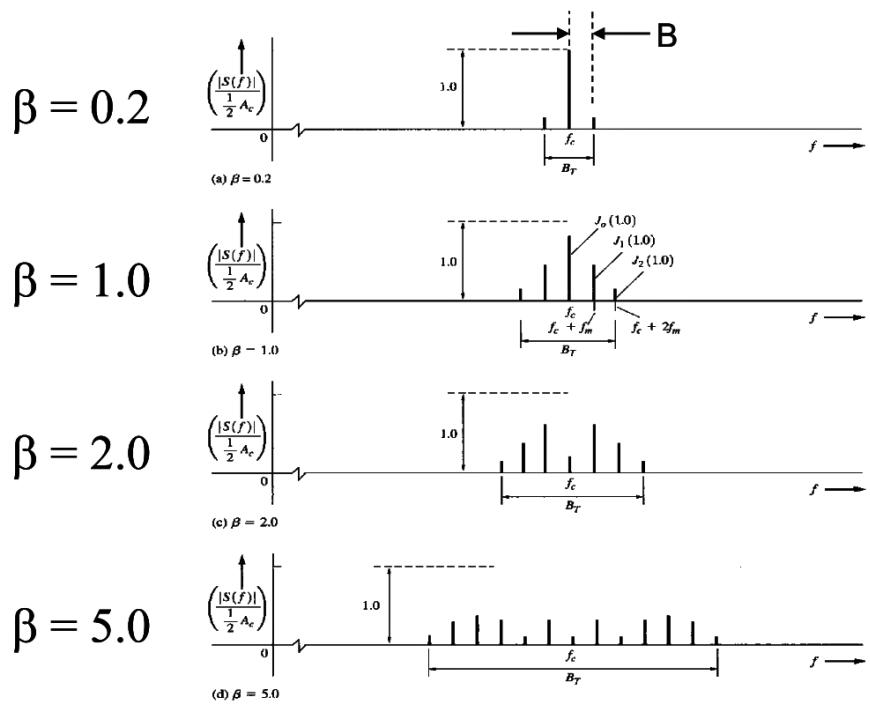


Figure 92: Frequency domain representation of an FM/PM signal, illustrating the impact of  $\beta$  on the bandwidth.

## 10.9 Instruction Exercises - AM/FM/PM

The solutions to these exercises may be found under the page **5ETC0 Canvas Page Modules** → **Week 5** → **I. Amplitude and Phase/Frequency modulation**

### Problem 19

Show that if  $v(t) = \operatorname{Re}\{g(t) \exp j\omega_c t\}$ , the equations below are correct, where  $g(t) = x(t) + jy(t) = R(t)e^{j\theta(t)}$

$$v(t) = R(t) \cos(\omega_c t + \theta(t))$$

$$v(t) = x(t) \cos(\omega_c t) - y(t) \sin(\omega_c t)$$

### Problem 20

A double-sideband suppressed carrier (DSB-SC) signal  $s(t)$  with a carrier frequency of 3.8 MHz has a complex envelope  $g(t) = A_c m(t)$ ,  $A_c = 50$  V, and the modulation is a 1-kHz sinusoidal test tone described by  $m(t) = 2 \sin(2\pi 1000t)$ . Evaluate the voltage spectrum for this DSB-SC signal.

### Exercise 21) (*Video solution available*)

Assume the DSC-SC voltage signal,

$$s(t) = 100 \sin(2\pi 1000t) \cos(2\pi 3.810 \cdot 10^6 t)$$

appears across a 50 Ohm resistive load

- Compute the actual average power dissipated the load.
- Compute the actual PEP.

### Exercise 22) (*Video solution available*)

A sine-shaped base-band signal  $m(t)$  is modulated by an amplitude modulator. The percentage of modulation of the AM signal  $s(t)$  at the output of the modulator is 50%.

- Give an expression for the signal  $s(t)$  at the output of the modulator and discuss the pros and cons of this modulation method.
- Draw the amplitude spectrum of the modulated signal and indicate clearly the relationship between the frequency components.
- Calculate the modulation efficiency  $E$  of the modulated signal.
- How can an AM signal with a percentage of modulation larger than 100% be demodulated?

### Exercise 23) (*Video solution available*)

$s(t) = 100 \cos((\omega_c - \omega_m)t) + 100 \cos(\omega_c t) + 100 \cos((\omega_c + \omega_m)t)$  Volts, is the voltage across a resistor of 50 ohms of a modulated signal.

- Show that  $s(t)$  is an amplitude modulated signal.
- Discuss the pros and cons of this modulation method.
- Draw a block diagram of an AM modulator.
- Calculate the modulation efficiency of the modulated signal.
- What is meant by the Peak Envelope Power and how large is the PEP in this case?
- In what way(s) can this signal be demodulated?

**Exercise 24) (*Video solution available*)**

A sinusoidal signal with a frequency of 1 kHz, is with a DSB-SC modulator modulated on a carrier with a frequency of 300 kHz, and loaded with a resistor of 50 Ohm. The amplitude of the DSB signal across the load resistance is 20 V.

- a) Give an expression for the DSB signal.
- b) Sketch the two-sided amplitude spectrum of the DSB signal (indicate axes, units and scales).
- c) Calculate the "Peak Envelope Power" that is dissipated in the load resistor.
- d) Calculate the percentage of modulation of the modulated signal.
- e) Draw the block diagram of an IQ (in-phase and quadrature-phase) detector.
- f) Derive expressions for the two outputs of the detector.

**Exercise 25) (*Video solution available*)**

The following two voltages across a load resistor of 50 ohms are given:

$$s_1(t) = 5\cos(2\pi 10^5 t)\sin(2\pi 10^9 t)$$

and

$$s_2(t) = 10 \sin(2\pi 10^9 t - 4 \cos(2\pi 10^5 t))$$

- a) From what type of modulation is each of those signals?
- b) What are the modulation indices or % modulation of these signals?
- c) Calculate the average power which is dissipated in the load resistor for each of the signals separately.
- d) What is, the bandwidth required for virtually distortion-free transfer of each of these signals?

**Exercise 26) (*Video solution available*)**

A modulating signal  $m(t) = m_0 \sin(\omega_m t)$  is AM modulated on a carrier (DSB with carrier).

From the AM signal the ratio of the maximum amplitude  $[A(t)]_{\max}$ , and the minimum amplitude  $[A(t)]_{\min}$  is equal to 2.

- a) Give a general expression for the complex envelope of an AM signal and give an expression for this AM signal.
- b) Calculate the percentage modulation.
- c) Why the percentage modulation usually has to be less than 100% in case of AM?

The AM signal is detected with an IQ (in-phase and quadrature-phase) detector. The phase of the local oscillator signal of the detector has a difference of 30 degrees with respect to the carrier of the AM signal.

**Exercise 27) (*Video solution available*)**

A baseband signal  $m(t)$  is multiplied with a carrier signal.

- a) What is the type of modulation of the resulting signal?
- b) What is the modulation efficiency of this type of modulation?

The signal is multiplied in a demodulator with an oscillator signal, and filtered by a low-pass filter. The oscillator signal has the same carrier frequency, but is not strictly synchronously

with respect to the carrier signal. The phase deviation of the oscillator signal with respect to the carrier signal is  $\phi$ .

- c)Derive an expression for the demodulated signal, and thus proof that the strength of the signal depends on the phase deviation.
- d)What has to be done with the signal  $s(t)$  in order to ensure that demodulation by means of an envelope detector is possible, and indicate the condition which has to be ensured.
- e)Explain how quadrature multiplexing works.

**Exercise 28) (*Video solution available*)**

An information signal  $0.4 \sin(\omega_m t)$ , is modulated on a carrier wave with an AM modulator (DSB with carrier).

- a)Give an expression for the AM signal and calculate the % modulation.
- b)Calculate the modulation or power efficiency of the AM signal.
- c>Show that the maximum achievable power efficiency in case of AM is equal to 50% if the % modulation may not be larger than 100%.

The % modulation is increased to 150%. Subsequently, the AM signal is detected by an envelope detector with an ideal diode.

- d)Draw as accurately as possible, the output signal of the envelope detector.
- e)How can the AM signal be demodulated free of distortion and discuss the requirements to be met.

**Exercise 29) (*Video solution available*)**

A carrier wave with an amplitude of 20 volts and a frequency of 1 MHz is frequency-modulated with a sinusoidal signal with an amplitude of 4 V and a frequency of 2 kHz. The modulation index of the modulated signal is 2. (See table next page)

- a)Draw the amplitude spectrum of the FM signal.
- b)Calculate the frequency deviation constant of the modulator.
- c)Calculate the peak frequency deviation of the FM signal.

The FM signal is filtered by an ideal band-pass filter with a central frequency of 1 MHz and a bandwidth of 10 kHz.

- d)Calculate the percentage of the signal power that is transferred, and judge whether this is sufficient.

**Problem 30 was discarded.**

**Exercise 31) (*Video solution available*)**

A sinusoidal signal with a frequency of 5 kHz is modulated with an AM modulator on a carrier wave with a frequency of 820 kHz and this modulated carrier wave is loaded with a resistance of 50 Ohm. The power of the carrier wave that is dissipated in the load resistor is 5 W. The modulation efficiency of the AM signal is 20%.

- a)Calculate the amplitude of the carrier wave.
- b)Calculate the % modulation of the AM signal.
- c)Sketch the two-sided amplitude spectrum of the AM signal (indicate axes, units and scales).
- d)Which form of modulation is obtained if the lower sideband of the AM signal is rotated 180 degrees in phase?
- e)How can a Vestigial Side Band modulated signal be generated and what are

the advantages of this modulation method?

n	$\beta$	0.2	0.5	1	2	2.5	5	8	16	32
0		0.990	0.938	0.765	0.224	-0.048	-0.178	0.172	-0.175	0.138
1		0.100	0.242	0.440	0.577	0.497	-0.328	0.235	0.090	-0.027
2		0.005	0.031	0.115	0.353	0.446	0.047	-0.113	0.186	-0.140
3				0.020	0.129	0.217	0.365	-0.291	-0.044	0.009



4			0.002	0.034	0.074	0.391	-0.105	-0.203	0.142
5				0.007	0.020	0.261	0.186	-0.058	0.026
6					0.004	0.131	0.338	0.167	-0.133
7					0.001	0.053	0.321	0.183	-0.076
8						0.018	0.223	-0.007	0.100
9							0.126	-0.190	0.126
10							0.061	-0.206	-0.029

# 11 The Physical channel

## 11.1 Learning objectives

Students completing this chapter should have learned:

1. Can name at least 3 differences between a wired and wireless channel.
2. Can calculate the propagation losses of a wireless signal at various distances with or without the presence of a knife-edge obstacle.
3. Can calculate the drop in power in an optical fiber based on the wavelength and attenuation parameters of the fiber.
4. Understand the concept of light guiding in an optical fiber and the difference between single and multi-mode fibers.
5. Calculate the impact of mode dispersion in a multi-mode fiber on eventual maximum data transmission speeds.

## 11.2 Motivation

Regardless of the chosen encoding and digitization process, a critical aspect of every communication system is the communication channel. We distinguish in general between two variations when discussing the channel, the wired and the wireless channels. In the following sections we will describe in much detail the unique properties of the radio frequency (RF) wireless channel (*supporting common communication networks such as Wifi and mobile phones*) and the wired optical fiber channel (*The backbone network supporting global, and more and more local, transport of data between fixed sites*).

## 11.3 Wired and wireless comparison

Before we detail the specifics of these two specific cases we give below a table comparing the main attributes and differences between the two modes of data transport. They are not given in any specific order of importance but are included to illustrate how fundamentally different the two modes of data transport are and why we need to develop a separate mathematical and physical understanding of them.

Stability	Physical link is a stable medium with well-defined time-invariant properties	Transmission medium is dynamically changing due to user mobility and multipath propagation
Capacity	Capacity increase is accomplished by adding another cable/fiber or adding more frequencies/wavelengths on the cable/fiber used	Capacity increase is based on more sophisticated transceivers and smaller cell sizes, allowing for spectral re-use since available spectrum is limited
Reach	Maximum un-repeated link range is limited by attenuation and propagation distortions (mostly in optical fibers)	Link range is limited by both the transmission medium (attenuation, fading, distortion) as well as by spectral efficiency requirements
Cross talk	Interference and cross talk from other users either do not exist or can be computed in advance and dealt with since they are time invariant	Interference and cross talk from other users is inherent in the operation principle of cellular communications and are also time variant
Delay in channel	Delay in the transmission process is constant and length dependent	Delay is distance dependent and since the mobile station is moving, is constantly changing
Bit error rate	The Bit-Error-Rate (BER) is practically exponentially dependent on the SNR, meaning a small decrease or increase of signal power can greatly effect the amounts of errors detected	For simple systems the BER is only linearly dependent on SNR (and is generally poorer than wired links). This means that changing the signal power has little effect on error rate and better error performance should be obtained through signal processing
Fidelity	Since it is based on well behaved transmission media, transmission quality of wire (fiber) links is high	Since the transmission media is difficult to control and predict, transmission quality is usually low, unless special measures are used
Jamming and interception	Jamming and interception (eaves dropping) are almost impossible for optical fiber communications without physically disrupting the optical signals	Jamming a wireless signal is straightforward. Interception can also be accomplished very easily and it is therefore that encryption should be used to protect data and voice carried over a wireless link
Mobility	Establishing a link is location based. A link is always connecting two terminals/nodes regardless of the person which is connected to that node	A link is usually based on mobile equipment which is the possession of a specific person, and is thus not related to fixed location
Size	While they should be easy to handle, the end terminals can be of different size and weight as they are stationary	Mobile handsets must be limited in weight and size which limits battery size and antenna layout and design
Power	Power is provided from the power grid to both end terminals in the link, and until recently the amount of power consumed by the device was of little concern to the designers of the systems	Mobile handsets use batteries for power and efficient use of the limited power stored in these batteries is a major design consideration and limitation in establishing a wireless link
Health considerations (radiation/power)	Emitted power can be increased even beyond safe exposure levels (for example laser power in a fiber) since the system is closed to human intervention	Emitted output power from mobile handsets, as well as from mobile base stations must conform to strict environmental requirements to lower the risk of potential health hazards

Table 2: Comparison between wired and wireless channels

## 11.4 Wireless channel

### 11.4.1 RF propagation

The basic rule for calculating the power distribution of a wireless transmission system assumes power propagates in all directions equally on the surface of a sphere whose center is at the transmitter site.

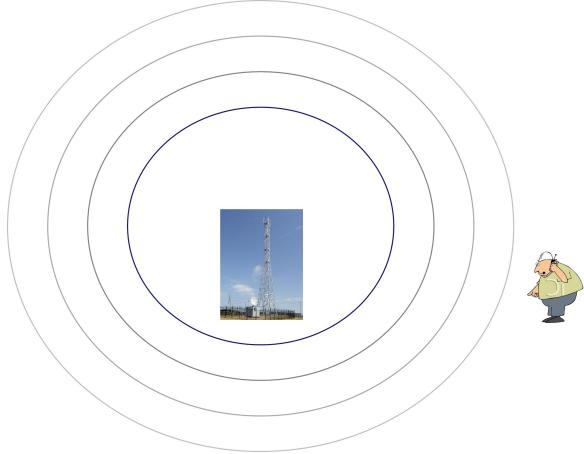


Figure 93: Energy is radiating in concentric spheres from the antenna to reach a user

In addition to the propagation losses, RF radiation is prone to atmospheric absorption. The amount of absorption is a function of the frequency and the humidity as can be seen in Figure 94.

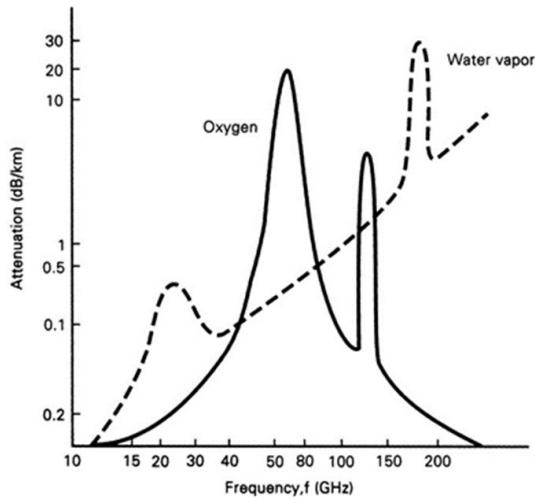


Figure 94: Attenuation of Radio waves as a function of frequency and humidity levels

### 11.4.2 Free space wireless propagation

The most simplified model for the energy to spread from the transmitter (in free space) is to assume a point transmitter emitting equal power in all directions. At a distance  $d$  one can imagine the power being equally distributed on the surface of a sphere with a surface area equal to  $4\pi d^2$ . The amount of power captured by the receiver can be simply calculated

by asking how large is the size (area) of the receiver antenna  $A_{Rx}$  compared with the total area of the sphere. Hence we obtain Equation 159:

$$P_{Rx}(d) = P_{Tx}G_{Tx}\frac{1}{4\pi d^2}A_{Rx} \quad (159)$$

For a given antenna, one can define the antenna gain, the ratio between the amount of power collected in the desired direction to a similar antenna with equal gain in all direction (*omnidirectional antenna*), using the formula below:

$$G_{Rx} = \frac{4\pi}{\lambda^2} A_{Rx} \quad (160)$$

Which when substituting into Equation 159 gives us **Frii's equation**.

$$P_{Rx}(d) = P_{Tx}G_{Tx}G_{Rx}\left(\frac{\lambda}{4\pi d}\right)^2 \quad (161)$$

Applying the logarithmic scaling to calculate the power in dB we can write:

$$P_{Rx}(d)[dBm] = P_{Tx} + G_{Tx} + G_{Rx} + 20\log_{10}\left(\frac{\lambda}{4\pi d}\right) \quad (162)$$

### **Example**

If the power of the transmitter is 0dBm and both transmitter and receive gains are 0dB, what would be the power at a distance of 1000 meters from the antenna if the  $\lambda = 4\pi$  [meter]?

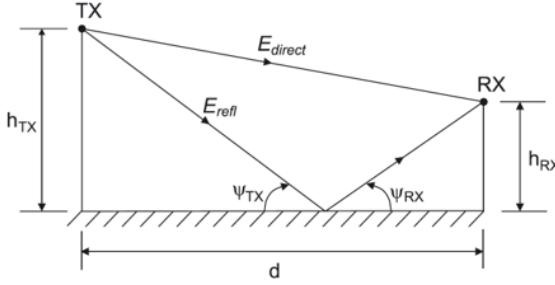
We can simply substitute the distance d into the equation and find that the power is -60dBm. And what if the distance now doubles?

#### **11.4.3 The case of a single reflection (Deriving the $d^{-4}$ rule)**

In reality, free space propagation is often not achieved in terrestrial systems. In most case, the signal propagating from the transmit to the receive antenna, will be reflected and refracted from adjacent building and ground surfaces. We analyze below a simple case of a single reflection from the ground. The scope of this course does not allow us to discuss in detail the magnitude and phase of electromagnetic waves reflecting from surfaces but this video illustrates that at very low grazing angles the reflected power is almost equal to the incident power ([https://www.youtube.com/watch?v=Ck0wvrBzu9I&t=4s&ab\\_channel=GetLearntw%2FChunck](https://www.youtube.com/watch?v=Ck0wvrBzu9I&t=4s&ab_channel=GetLearntw%2FChunck)).

As a result of the very strong reflection from the ground, and the fact that the direct signal and the reflected signal travel different paths, a phase difference is present. Figure 95 illustrates the geometry of the problem that will be analyzed with the equations below.

The derivation of the expected behavior of the total detected electric field at the receiver antenna is elaborated in the equations below. We start by writing an expression for the field strengths for the direct and deflected components as they are to be detected at the receiving antenna (Eqn. 163 and 165). For every field, we need to assume a different distance as calculated in Equations (164 and 166). Summing up the two fields we can write equations 167 and 168 and assuming the distance  $d$  is much larger than the height of either transmitter or receiver, we can write an approximate expression for the length difference between the two paths (Equation 169). If we substitute this difference into equation 168 and redefine the phase difference  $\Delta\phi$  based on the geometry of the problem, we conclude that for values of  $d$  which are far larger than  $(h_{Tx}h_{Rx})/\lambda$  we can write the total field strength using euqation 172.



**Figure 4.15** Geometry for the derivation of the  $d^{-4}$  law.

Figure 95: Radio propagation with a single reflection from the ground.

$$E_{direct}(d_{direct}) = E(1m) \left( \frac{1}{d_{direct}|m|} \right) \exp[j(2\pi f_c t - 2\pi f_c \frac{d_{direct}}{c_0})] \quad (163)$$

$$d_{direct} = \sqrt{(h_{TX} - h_{RX})^2 + d^2} \quad (164)$$

$$E_{refl}(d_{refl}) = (-1)E(1m) \left( \frac{1}{d_{refl}|m|} \right) \exp[j(2\pi f_c t - 2\pi f_c \frac{d_{refl}}{c_0})] \quad (165)$$

$$d_{refl} = \sqrt{(h_{TX} + h_{RX})^2 + d^2} \quad (166)$$

$$E_{tot}(d) = E(1m) \left( \frac{1}{d[m]} \right) \{ \exp[j(2\pi f_c t - 2\pi f_c \frac{d_{direct}}{c_0})] - \exp[j(2\pi f_c t - 2\pi f_c \frac{d_{refl}}{c_0})] \} \quad (167)$$

$$E_{tot}(d) = E(1m) \left( \frac{1}{d[m]} \right) \exp[j(2\pi f_c t - 2\pi f_c \frac{d_{direct}}{c_0})] \{ 1 - \exp[-j(2\pi f_c \frac{d_{refl} - d_{direct}}{c_0})] \} \quad (168)$$

$$d_{refl} - d_{direct} = 2 \frac{h_{Tx} h_{Rx}}{d} \quad (169)$$

$$|E_{tot}(d)| \approx E(1m) \frac{1}{d[m]} \sqrt{(1 - \cos(\Delta\varphi))^2 + \sin^2(\Delta\varphi)} \quad (170)$$

$$\Delta\varphi = 2 \frac{h_{Tx} h_{Rx}}{d} \frac{2\pi f_c}{c_0} \quad (171)$$

$$\text{In the case when } d_{limit} \gg \frac{h_{Tx} h_{Rx}}{\lambda}, |E_{tot}(d)| \approx E(1m) \frac{4\pi h_{Tx} h_{Rx}}{\lambda d^2} \quad (172)$$

We define a distance  $d_{break}$  as the distance where the propagation attenuation is dominated by the multipath interference

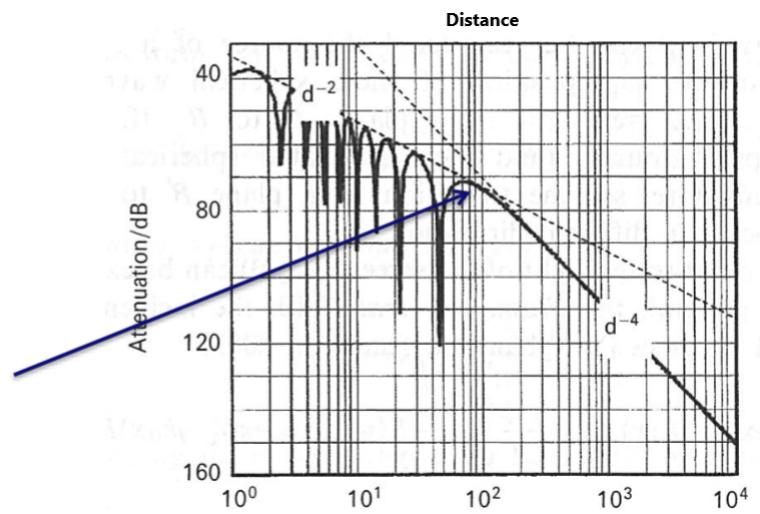
$$d_{break} = \frac{4\pi h_{Tx} h_{Rx}}{\lambda} \quad (173)$$

Which means that assuming the power is proportional to the square of the electric field, we can write an approximated value for the power at the receiver as:

$$P_{Rx}(d) \approx P_{Tx} G_{Tx} G_{Rx} \left( \frac{h_{Tx} h_{Rx}}{d^2} \right)^2 \quad (174)$$

Comparing this equation to Frii's equation (Equation 161) we see that the power will drop off with the 4th power of the distance, if  $d > d_{break}$ . This is also visible when looking at the curve of the power detected as a function of distance which exhibits two distinct slopes, one of  $1/d^2$  and the other of  $1/d^4$ .

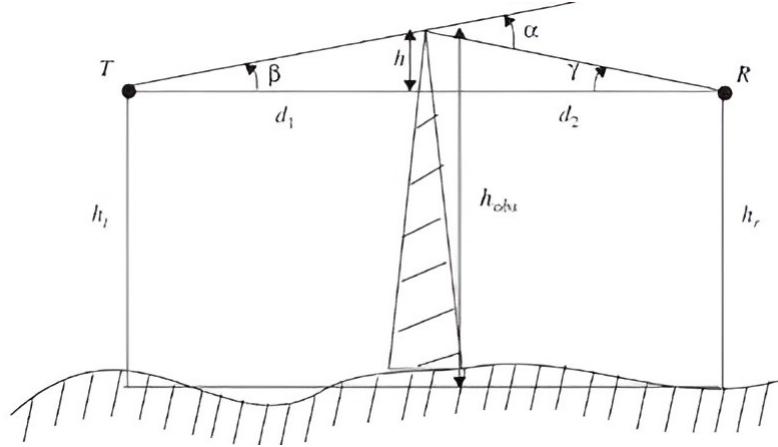
$$d_{break} = \frac{4\pi h_{Tx} h_{Rx}}{\lambda}$$



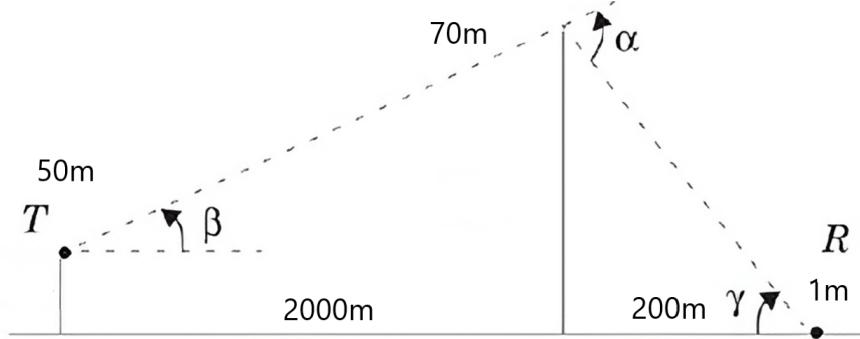
#### 11.4.4 Knife Edge

To analyze diffraction and other effects in relation to wireless propagation, we can take a look at the "Knife Edge" style situations.

These are the case where an object is blocking the RF path, as in the depiction below.



Consider the following details, and the fact that the transmission power is equal to 1W. Moreover the gain of TX and RX are 1 (0 dB).



Let's imagine that we are operating at a frequency of 1GHz, and we are tasked with computing the received power.

Firstly, we will need to compute  $\alpha$ , in the following manner (using pre-knowledge trig relations).

$$\alpha = \beta + \gamma \quad (175)$$

$$= \arctan\left(\frac{h_{obs} - h_{Tx}}{d_1}\right) + \arctan\left(\frac{h_{obs} - h_{Rx}}{d_2}\right) \quad (176)$$

$$= \arctan\left(\frac{70 - 50}{2000}\right) + \arctan\left(\frac{70 - 1}{200}\right) \quad (177)$$

$$= 0.3422 \quad (178)$$

Since we are given the operating frequency, we compute the wavelength  $\lambda = \frac{c}{f} = 0.3m$ .

Next we can compute  $\nu$ ,

$$\nu = \alpha \sqrt{\frac{2d_1 d_2}{\lambda(d_1 + d_2)}} \quad (179)$$

$$= 0.3422 \sqrt{\frac{2 \cdot 2000 \cdot 200}{0.3(2000 + 200)}} \quad (180)$$

$$= 11.9138 \quad (181)$$

This allows to compute the power loss,

$$A(v) = P_{loss} = 6.9 + 20 \log_{10}\{\sqrt{v^2 + 1} + v - 0.1\} \quad (182)$$

$$= 6.9 + 20 \log_{10}\{\sqrt{11.9138^2 + 1} + 11.9138 - 0.1\} \quad (183)$$

$$= 34.42 \text{ dB} \quad (184)$$

Before continuing, we must compute the  $d_{break}$ , given by

$$d_{break} = \frac{4\pi h_{Tx} h_{Rx}}{\lambda} = \frac{4 \cdot \pi \cdot 50 \cdot 1}{0.3} = 2094.4m \quad (185)$$

As mentioned before, we must use the equation which corresponds either to the case of  $d_{total} < d_{break}$  or  $d_{total} > d_{break}$ .

Since in this case, the total distance between TX and RX is above  $d_{break}$ , the following equation is used:

$$P_{Rx}(d) \approx P_{Tx} G_{Tx} G_{Rx} \left(\frac{h_{Tx} h_{Rx}}{d^2}\right)^2 \quad (186)$$

$$= P_{Tx} G_{Tx} G_{Rx} \left(\frac{50 \cdot 1}{(2000 + 200)^2}\right)^2 \quad (187)$$

$$= 1.0672 \cdot 10^{-10} W = -99.7175 \text{ dBW}. \quad (188)$$

Thus we reach the final calculation, based on the free-space loss and the added loss from the diffraction.

$$P_{RX} = P_{Rx}(d) - P_{loss} = -99.7175 - 34.42 = -134.1375 \text{ dBW} \quad (189)$$

**As a practice exercise, attempt the problem for an operating frequency of 2GHz instead. Hint: recompute  $d_{break}$ .**

The alternative equation (for the  $d_{total} < d_{break}$  scenario) is

$$P_{Rx}(d) = P_{Tx} G_{Tx} G_{Rx} \left(\frac{\lambda}{4\pi d}\right)^2 \quad (190)$$

## 11.5 Wired Connections

### 11.5.1 Introduction

As discussed in the comparison section 11.3 wired connections (channels) exhibit very different properties when compared to wireless channels (*Try to recall at least three differences without going back to that section, did you succeed?*). Early days of wired connection were based on the Telegraph and later telephone lines. They were based on the use of metal wires (mostly copper as a trade-off between high conductivity and price) physically connecting locations.

For more than a century copper-based wired connections provided an excellent base for a global network of telecommunications (signaling and voice). However as more and more services became digital and (as explained in earlier sections) the amount of data generated increased, the speed at which data was required to be transported on wired lines grew. Metal conductors provide a good means of transport of information through the propagation of charge (electrons in the form of current).

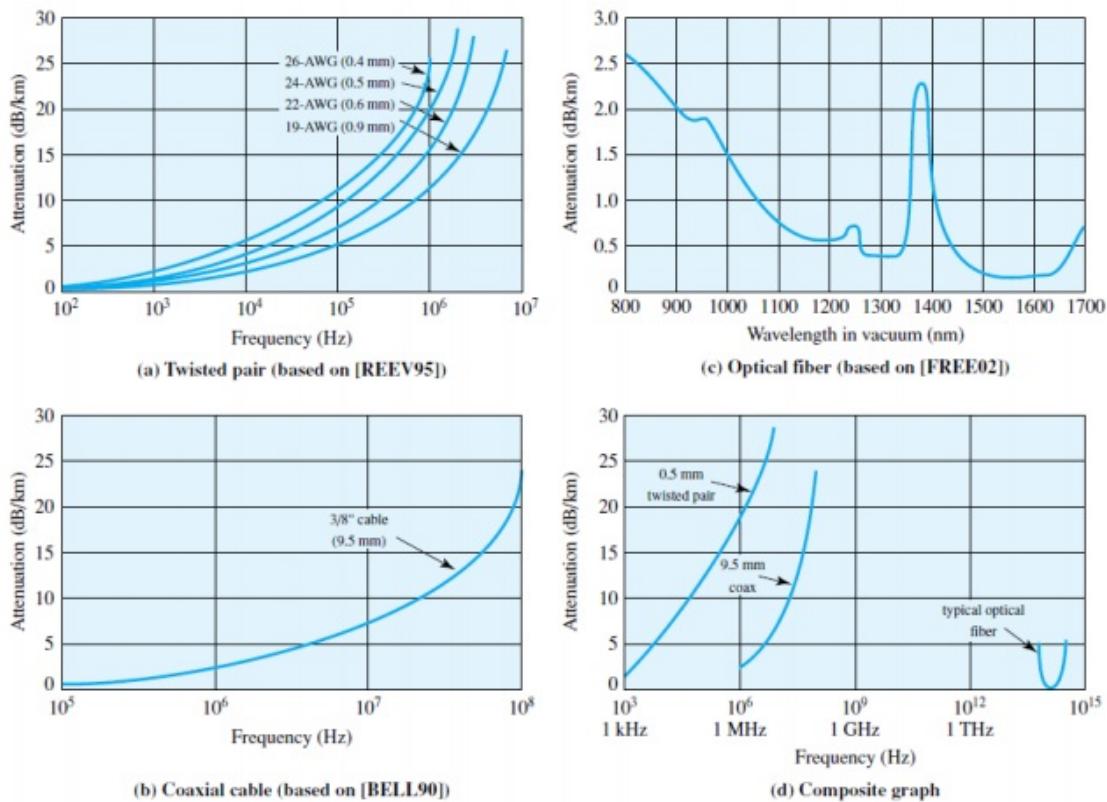


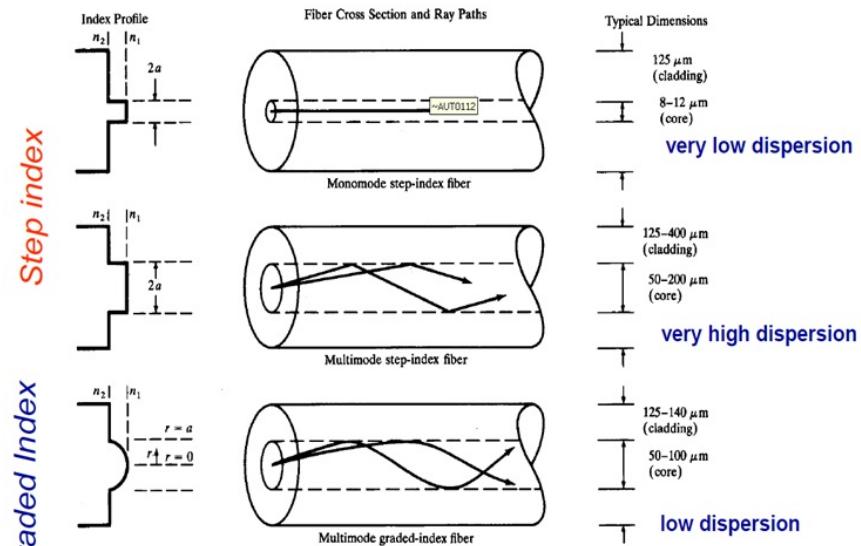
Figure 96: Losses of different kinds of cables as a function of frequency (wavelength)

In figure 96 we can see that for twisted pair and coaxial cables, the attenuation grows very quickly (note that the attenuation values are in dB so the loss is growing even faster than exponential growth!) with increasing frequency. In contrast, for optical fibers, we see the loss at optical wavelengths (or frequencies) the loss at around 1000nm is only 1dB per kilometre. Furthermore, this loss does not seem to be dependent on the frequency of the information carried.

**Question** If the wavelength is 1500nm, what is the frequency of the wave in vacuum

(n=1)?

Optical fibers (the use of thin strands of glass as a means of transporting data) were first proposed by Charles Kao in the 1960's and have been perfected to provide ultra-low loss in the following years. First generation of optical fibers were based on a relatively thick glass core, with either a step (abrupt) change in refractive index or a graded (gradual) change in refractive index (see Figure 97). Early optical fiber systems, were running at relatively low symbol rates. Hence, large core fibers, which supported multiple propagation modes (illustrated by different angles in the fiber in Figure 97), were perfectly suitable.



**FIGURE 2-10**  
Comparison of single-mode and multimode step-index and graded-index optical fibers.

Figure 97: Illustration of two rays propagating in a glass waveguide of thickness  $d$  and refractive indices  $n_1$  and  $n_2$

We can better understand the concept of wave propagation in multi-mode fibers but solving the problem of co-propagation of two rays in a waveguide of thickness  $d$  and refractive indices  $n_1$  and  $n_2$  for the core and cladding respectively. In order for the wave, represented by the two rays to continue to propagate, we require that the phase accumulated by ray 1 from point A to point B is equal to the phase ray 2 accumulates when traversing the distance C to D (see Figure 98).

The complete derivation can be found here: <https://www.ecb.torontomu.ca/~courses/ee8114/projects03/klavir/evgenyklavir.pdf>. Following this analysis (which theory is beyond the scope of this course!) we obtain Equation 11.5.1. This formula cannot be solved analytically, but by comparing the right and left side of this equation numerically we can find values of the angle  $\theta$  which satisfy the equality. We define these values as supported angles or propagation.

$$\tan\left(\frac{\pi d \sin(\theta)}{\lambda} - \frac{\pi m}{2}\right) = \frac{\sqrt{\cos^2(\theta) - (n_1/n_2)^2}}{\sin(\theta)} \quad (191)$$

In order to better understand the limitations of multi-mode fiber transmission please check the minilab on the next page.

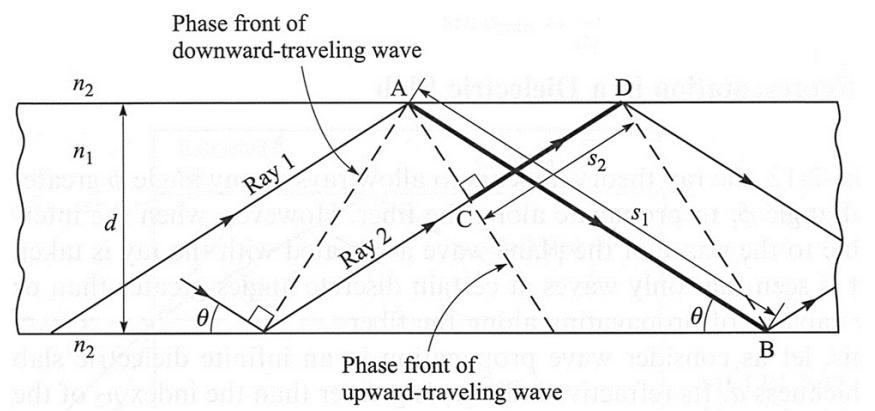
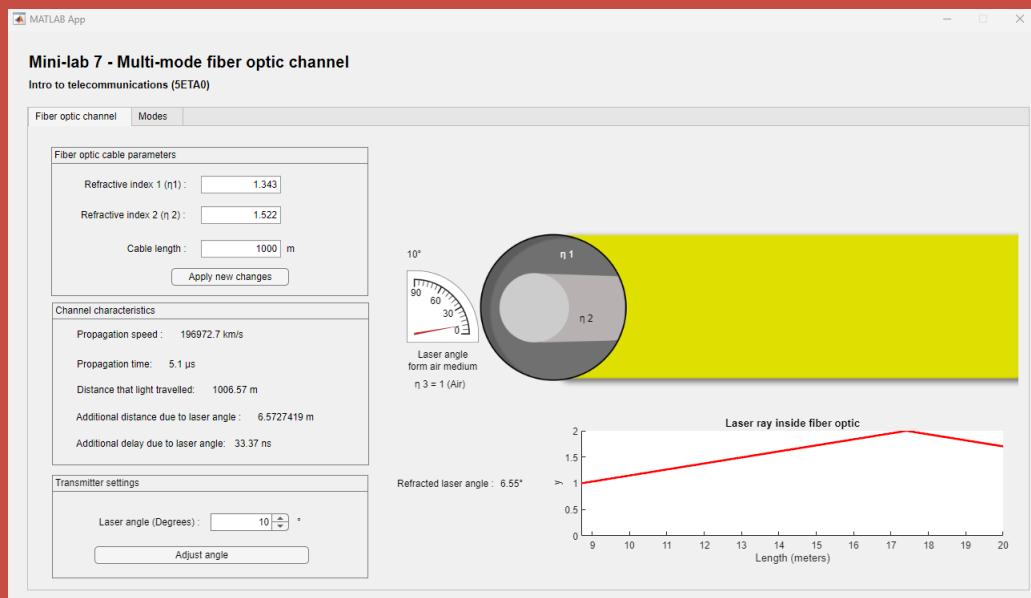


Figure 98: Tracing two rays in a wave guide of thickness  $d$  to find the condition for mode propagation

## Minilab exercise 12.1

This mini-lab exercise requires you to use Mini-lab 7 - Multi-mode fiber optic channel



In this minilab, you can change various settings of the fiber optic channel, such as the laser angle, the refractive index of the core and cladding of the fiber optic. The purpose of this minilab is to show how different angles influence the propagation time of the rays, total internal reflection, and if the fiber optic where a single mode, the difficulty on getting the right angle, for the single mode to propagate. Any changes you make, you have to click on 'Apply new changes', and if you change the laser angle click on 'adjust angle'

- Open the minilab, on the default configuration. Keep increasing the laser angle. What happens to the propagation time? Moreover, after a certain angle there is no TIR (total internal reflection), can you figure out this angle? Try to calculate the values you see on the minilab in order to help you better understand.
- So far you have been experimenting on a multi-mode fiber optic cable, now to see the difficulties of the single mode, you may go to the tab named '*Modes and fiber settings*' and there you can change settings such as the laser light wavelength, the core diameter of the fiber optic. If you click plot changes, you may see the modes equation getting plotted, try to pick an angle that satisfies the condition for rays to propagate on this fiber optic cable. On the fiber optic channel tab, you may check if the ray can propagate in a single-mode fiber, given the laser angle, wavelength and core diameter. With the default configuration can you find a suitable laser angle that can make the rays propagate?

## 11.6 Exercises

### Question 33

Given the following geometry determine:

- The added loss due to the knife-edge diffraction,
- The height of the obstacle for the case of 6 dB diffraction loss (assume  $f=800$  MHz)

### Question 35

We want to send a signal over 50km from point A to point B in a fiber link. What is the power received at point B?

Use:

- Launched power 0dBm
- Fiber loss – 0.25dB/km
- Gain Tx and Rx =1
- Optical frequency – 193THz

### Question 36

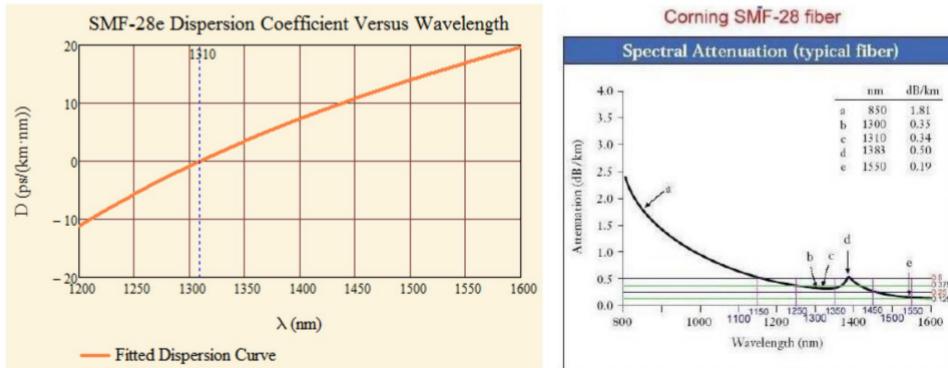
An internet router connected to a data center with a multi-mode fiber of 10km which supports two modes (0 and 2 deg).

- What is the delay between the two fiber modes? Assume that the refractive index of the glass in the fiber is equal to 1.5.
- Let the RMS delay spread for a multi-path wireless communication be 37.3 ns (derived in Question 34). Which communication is faster between the two cases if the delay between the two fiber modes can be considered as the RMS delay spread of the optical fiber link?

### Question 37

An optical fiber is drawn between a receiver and a transmitter. The sketches of the optical attenuation as a function of wavelength and the fiber dispersion as a function of wavelength are shown below.

- Discuss and motivate which wavelength should be preferred ( $1.55 \mu\text{m}$  or  $1.31 \mu\text{m}$ ) for the case of transmitting a signal with 10 MHz bandwidth.
- Would your choice be different if the optical carrier must be modulated at 100Gb/sec? Explain!!



## 12 Appendix

### 12.1 Working with dB's

Irrespective of what field you will follow further it is highly probable that you will need to work with Decibel units. Especially with regards to antenna systems or communications, they allow us to quickly grasp the scale of a certain power or noise level.

For example, it is easier to compare 76dB with 102dB, rather than their linear counterparts of 39810717 and 15848931922...

Since the use of Decibel is simply based on the shift from a linear to a logarithmic scale we can perform this conversion for any unit we want. In textbooks you can hence find  $dB_V$  for example which is a conversion of voltage into a logarithmic scale. In this course we will only discuss the conversion of power into logarithmic scale hence we will discuss  $dB_W$  or  $dB_m(\text{watt})$ . These units represent respectively, the conversion of power in Watts and milliWatts into a logarithmic scale.

When working with power or power ratio's we will use the following forumli to convert between the linear and dB scale:

$$P_{dBm} = 10 \cdot \log_{10} P_{mWatt}$$

or:

$$P_{dBW} = 10 \cdot \log_{10} P_{Watt}$$

If power is given in  $dBm$  or  $dBW$  one can use the equations below to convert to milliWatts or Watts respectively:

$$P_{mWatt(Watt)} = 10^{\frac{P_{dBm(W)}}{10}}$$

For values such as Gain or Signal to Noise ratio which are unitless, we use the same equations but replace powers with ratios. For example, if the  $SNR = \frac{P_{signal}}{P_{noise}}$  is given as a linear relationship one can convert it into dB by using:

$$SNR_{dB} = 10 \cdot \log 10 SNR_{linear}$$

When working with logarithmic values one needs to be careful to recall the properties of working with the log function. Recall that:

$$\log(A * B) = \log(A) + \log(B)$$

and that:

$$\log\left(\frac{A}{B}\right) = \log A - \log B$$

It is hence possible to calculate the power after the attenuation of a free-space channel by writing:

$$P_{Rx} = P_{Tx} - L$$

(Where  $L$  is the channel loss in dB)

This is equivalent to writing:

$$P_{Rx} = \frac{P_{Tx}}{L}$$

(Where we have L is an attenuation ratio)

The use of decibels also allows us to quickly estimate the power ratio between two magnitude as we can use the following thumb rules for conversion:

Ratio	dB
2	3
3	5
4	6
5	7
6	8
8	9

For further practice, I suggest checking out: <https://greatscottgadgets.com/sdr/3/>.

## 13 Physical channel equations

$$P_{Rx}(d) = P_{Tx} \frac{1}{4\pi d^2} A_{Rx} \quad (192)$$

$$G_{Rx} = \frac{4\pi}{\lambda^2} A_{Rx} \quad (193)$$

$$P_{Rx}(d) = P_{Tx} G_{Tx} G_{Rx} \left( \frac{\lambda}{4\pi d} \right)^2 \quad (194)$$

$$\frac{\sin(\theta_t)}{\sin(\theta_e)} = \frac{\sqrt{real(\delta_1)}}{\sqrt{real(\delta_2)}} = \frac{\sqrt{\epsilon_1}}{\sqrt{\epsilon_2}} \quad (195)$$

$$P_{Rx}(d) \approx P_{Tx} G_{Tx} G_{Rx} \left( \frac{h_{Tx} h_{Rx}}{d^2} \right)^2 \quad (196)$$

$$d_{break} = \frac{4\pi h_{Tx} h_{Rx}}{\lambda} \quad (197)$$

$$\nu = \alpha \sqrt{\frac{2d_1 d_2}{\lambda(d_1 + d_2)}} \quad (198)$$

$$\alpha = \beta + \gamma \quad (199)$$

$$\beta = \arctan\left(\frac{h_{obs} - h_{Tx}}{d_1}\right) \quad (200)$$

$$\gamma = \arctan\left(\frac{h_{obs} - h_{Rx}}{d_2}\right) \quad (201)$$

$$A(v) = 6.9 + 20 \log_{10} \{ \sqrt{v^2 + 1} + v - 0.1 \} \quad (202)$$

$$\bar{\tau} = \frac{\sum_k P(\tau_k) \tau_k}{\sum_k P(\tau_k)} \quad (203)$$

$$\sigma_{tau} = \sqrt{Avg(\tau^2) - (\bar{\tau})^2} \quad (204)$$

$$|r(t)|^2 = \left| \sum_{i=0}^{N-1} a_i \exp(j\theta_i(t, \tau)) \right|^2 \quad (205)$$

$$\Delta\phi = \frac{2\pi\Delta l}{\lambda} = \frac{2\pi\nu\Delta t}{\lambda} \cos(\theta) \quad (206)$$

$$f_d = \frac{1}{2\pi} \frac{\Delta\phi}{\Delta t} = \frac{\nu}{\lambda} \cos(\theta) \quad (207)$$

## 14 Appendix C: Autocorrelation and PSD relation

**Step 1: Definition of the autocorrelation function  $R_{ww}(\tau)$**

$$R_{ww}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} w(t) w(t + \tau) dt. \quad (208)$$

This expression calculates how similar the signal  $w(t)$  is to a delayed version of itself  $w(t + \tau)$ , averaged over an interval of length  $T$ . As  $T$  goes to infinity, it represents the long-term average.

**Step 2: Taking the Fourier transform of  $R_{ww}(\tau)$**

$$F[R_{ww}(\tau)] = \int_{-\infty}^{\infty} R_{ww}(\tau) e^{-j\omega\tau} d\tau. \quad (209)$$

Here, the Fourier transform converts the autocorrelation function  $R_{ww}(\tau)$  from the time (delay) domain into the frequency domain.

**Step 3: Substituting the definition of  $R_{ww}(\tau)$  into the transform**

$$F[R_{ww}(\tau)] = \int_{-\infty}^{\infty} \left( \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} w(t) w(t + \tau) dt \right) e^{-j\omega\tau} d\tau. \quad (210)$$

The limit  $\lim_{T \rightarrow \infty}$  appears because the true autocorrelation function is defined in the limit of large  $T$ . Inside, there is a product  $w(t)w(t + \tau)$  integrated over  $-T/2$  to  $T/2$ .

**Step 4: Interchanging limits and integrals (assuming conditions for valid interchange)**

$$F[R_{ww}(\tau)] = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} \int_{-\infty}^{\infty} w(t) w(t + \tau) e^{-j\omega\tau} d\tau dt. \quad (211)$$

Under certain regularity conditions (such as absolute integrability), it is possible to exchange the order of integration and the limit.

**Step 5: Noting that  $w(t)$  is real and introducing its Fourier transform**

$$w^*(t) = w(t) \quad (\text{since } w(t) \text{ is real}), \quad (212)$$

$$W(\omega) = \int_{-\infty}^{\infty} w(t) e^{-j\omega t} dt. \quad (213)$$

Because  $w(t)$  is real, its complex conjugate  $w^*(t)$  equals  $w(t)$ . The function  $W(\omega)$  is the Fourier transform of  $w(t)$ .

**Step 6: Relating to  $|W(f)|^2$**

$$F[R_{ww}(\tau)] = \lim_{T \rightarrow \infty} \frac{1}{T} |W(\omega)|^2. \quad (214)$$

In signal processing, it can be shown (using the convolution theorem and properties of the Fourier transform) that the transform of the autocorrelation function corresponds to  $|W(\omega)|^2$  scaled by  $\frac{1}{T}$ . When taking  $T \rightarrow \infty$ , this defines the power spectral density.

**Step 7: Defining the power spectral density  $P_w(f)$**

$$P_w(f) = \lim_{T \rightarrow \infty} \frac{1}{T} |W(f)|^2. \quad (215)$$

This final step states that the power spectral density  $P_w(f)$  is the limiting form of  $\frac{1}{T}|W(f)|^2$  as  $T$  approaches infinity. It shows how power is distributed over frequency for the signal  $w(t)$ .

By looking at the last two equations, the relation can be clearly made that  $P_w(f) = F[R_{ww}(\tau)]$ .

## References

- [1] <https://dspillustrations.com/pages/posts/misc/the-dirac-comb-and-its-fourier-transform.html>.
- [2] L. W. Couch, M. Kulkarni, and U. S. Acharya, *Digital and analog communication systems*. Citeseer, 2013, vol. 8.
- [3] A. Lyon, “Why are normal distributions normal?” *The British Journal for the Philosophy of Science*, 2014.